

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 21, 2001, 00:35:06 ; Search time 3479.37 Seconds
(without alignments)
97.802 Million cell updates/sec

Title: US-09-138-735-9

Perfect score: 22

Sequence: 1 cagccgacgtagctgcgtcct 22

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues 2688314

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

GenEmbl:*

- 1: gb_ba1:*
- 2: gb_ba2:*
- 3: gb_ba3:*
- 4: gb_in1:*
- 5: gb_in2:*
- 6: gb_in3:*
- 7: gb_om:*
- 8: gb_ov:*
- 9: gb_pat1:*
- 10: gb_pat2:*
- 11: gb_ph:*
- 12: gb_p11:*
- 13: gb_p12:*
- 14: gb_p13:*
- 15: gb_p14:*
- 16: em_ba1:*
- 17: em_ba2:*
- 18: em_fun:*
- 19: em_htgo_hum:*
- 20: em_htgo_inv:*
- 21: em_htgo_rod:*
- 22: em_htg_hum1:*
- 23: em_htg_hum2:*
- 24: em_htg_hum3:*
- 25: em_htg_hum4:*
- 26: em_htg_hum5:*
- 27: em_htg_hum6:*
- 28: em_htg_hum7:*
- 29: em_htg_hum8:*
- 30: em_htg_inv1:*
- 31: em_htg_inv2:*
- 32: em_htg_other:*
- 33: em_htg_rod:*
- 34: em_hum1:*
- 35: em_hum2:*
- 36: em_hum3:*
- 37: em_hum4:*
- 38: em_hum5:*
- 39: em_hum6:*
- 40: em_hum7:*
- 41: em_in:*
- 42: em_om:*
- 43: em_or:*

- 44: em_ov:*
- 45: em_pat:*
- 46: em_ph:*
- 47: em_pl:*
- 48: em_ro:*
- 49: em_sts:*
- 50: em_sy:*
- 51: em_un:*
- 52: em_vi:*
- 53: gb_sts1:*
- 54: gb_sts2:*
- 55: gb_sts3:*
- 56: gb_sy:*
- 57: gb_un:*
- 58: gb_vil:*
- 59: gb_vil2:*
- 60: gb_htg1:*
- 61: gb_htg2:*
- 62: gb_htg3:*
- 63: gb_htg4:*
- 64: gb_htg5:*
- 65: gb_htg6:*
- 66: gb_htg7:*
- 67: gb_htg8:*
- 68: gb_htg9:*
- 69: gb_htg10:*
- 70: gb_htg11:*
- 71: gb_htg12:*
- 72: gb_htg13:*
- 73: gb_htg14:*
- 74: gb_htg15:*
- 75: gb_htg16:*
- 76: gb_htg17:*
- 77: gb_htg18:*
- 78: gb_htg19:*
- 79: gb_htg20:*
- 80: gb_htg21:*
- 81: gb_htg22:*
- 82: gb_htg23:*
- 83: gb_htg24:*
- 84: gb_htg25:*
- 85: gb_pr1:*
- 86: gb_pr2:*
- 87: gb_pr3:*
- 88: gb_pr4:*
- 89: gb_pr5:*
- 90: gb_pr6:*
- 91: gb_pr7:*
- 92: gb_pr8:*
- 93: gb_pr9:*
- 94: gb_rot1:*
- 95: gb_rot2:*
- 96: gb_in4:*
- 97: gb_pr10:*
- 98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	22	100.0	22	9 A48918	A48918 Sequence 9
2	22	100.0	22	9 AR047927	AR047927 Sequence 9
3	22	100.0	3402	9 A48910	A48910 Sequence 1
4	22	100.0	3402	9 AR047920	AR047920 Sequence 1
5	22	100.0	3402	96 TC024190	TC024190 Sequence 96
6	17.8	80.9	158566	66 AC021683	AC021683 Homo sapi
7	17.2	78.2	29485	4 AC005439	AC005439 Drosophila
8	17.2	78.2	33077	64 AC017302	AC017302 Drosophila

9	17.2	78.2	36676	93	HSJ858B16	AL096768 Human DNA
10	17.2	78.2	97308	63	AC013925	AC013925 Drosophila
11	17.2	78.2	151369	13	AP001080	AP001080 Oryza sat
12	17.2	78.2	161046	58	AF081810	AF081810 Lymantria
13	17.2	78.2	168655	13	AP000969	AP000969 Oryza sat
14	17.2	78.2	175682	4	AC007417	AC007417 Drosophila
15	17.2	78.2	261846	5	AE003830	AE003830 Drosophila
16	16.8	76.4	1188	14	GB585	X9390 G. biloba DN
17	16.8	76.4	66915	63	AC013073	AC013073 Drosophila
18	16.8	76.4	117486	83	AP003037	AP003037 Homo sapi
19	16.8	76.4	127652	82	AP001526	AP001526 Homo sapi
20	16.8	76.4	152936	60	AC004229	AC004229 Homo sapi
21	16.8	76.4	222472	69	AC026091	AC026091 Homo sapi
22	16.4	74.5	11475	13	AE005794	AE005794 Caulobact
23	16.4	74.5	107600	13	ATF16J13	AL049638 Arabidops
24	16.4	74.5	110000	84	LMFICR32_15	Continuation (16 o
25	16.4	74.5	190026	13	ATCHRIV33	AL161533 Arabidops
26	16.2	73.6	2426	1	AF047430	AF047430 Stenothrips
27	16.2	73.6	2429	89	AK026936	AK026936 Homo sapi
28	16.2	73.6	4051	88	AF006005	AF006005 Homo sapi
29	16.2	73.6	5638	91	D83784	D83784 Human mRNA
30	16.2	73.6	10160	1	AE002024	AE002024 Halobacte
31	16.2	73.6	10282	1	AE005100	AE005100 Halobacte
32	16.2	73.6	10413	1	AE005018	AE005018 Halobacte
33	16.2	73.6	10475	10	I08101	I08101 Sequence 4
34	16.2	73.6	10475	10	I08101	I08101 Sequence 9
35	16.2	73.6	10476	97	HUMSPBAA	M24461 Human pulm
36	16.2	73.6	24008	64	AC017420	AC017420 Drosophila
37	16.2	73.6	28372	6	CET07A5	Z48055 Caenorhabdi
38	16.2	73.6	38397	6	DMC8D	AL022018 Drosophila
39	16.2	73.6	54327	78	AC090435	AC090435 Chlamydom
40	16.2	73.6	58198	78	AC091323	AC091323 Mus muscu
41	16.2	73.6	60135	78	AC091200	AC091200 Drosophila
42	16.2	73.6	62148	78	AF192524	AF192524 Homo sapi
43	16.2	73.6	88050	78	AC090436	AC090436 Chlamydom
44	16.2	73.6	104171	65	AC019499	AC019499 Drosophila
45	16.2	73.6	110098	90	AL358976	AL358976 Human DNA

ALIGNMENTS

RESULT 1

LOCUS A48918 22 bp DNA PAT 07-MAR-1997

DEFINITION Sequence 9 from Patent WO9605312.

ACCESSION A48918

VERSION A48918.1 GI:2302577

KEYWORDS

SOURCE unclassified.

ORGANISM unclassified.

REFERENCE 1 (bases 1 to 22)

AUTHORS Paranhos-Baccala, G., Lesenechal, M. and Jolivet, M.

TITLE NOVEL TRYPA NOSOMA CRUZI ANTIGEN, AND GENE CODING THEREFOR, THEIR APPLICATION TO THE DETECTION OF CHAGAS' DISEASE

JOURNAL Patent: WO 9605312-A 9 22-FEB-1996;

COMMENT BIO MERIEUX (FR)

Other publication CA 2173957 960222

Other publication AU 3169195 960307

Other publication FR 2723589 960216.

FEATURES

source 1..22

/organism="unclassified"

/db_xref="taxon:32644"

BASE COUNT 3 a 8 c 7 g 4 t

ORIGIN

Query Match 100.0%; Score 22; DB 9; Length 22;

Best Local Similarity 100.0%; Pred. No. 15;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cagccgacgtagctgcgtcct 22

|||||

Db 1 CAGCCGACGCTAGCTGCTCT 22

RESULT 2

LOCUS AR047927 22 bp DNA PAT 29-SEP-1999

DEFINITION Sequence 9 from patent US 5820864.

ACCESSION AR047927

VERSION AR047927.1 GI:5970270

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 22)

AUTHORS Paranhos-Baccala, G., Lesenechal, M. and Jolivet, M.

TITLE TRYPA NOSOMA CRUZI ANTIGEN, gene encoding therefor and methods of detecting and treating chagas disease

JOURNAL Patent: US 5820864-A 9 13-OCT-1998;

FEATURES

source 1..22

/organism="unknown"

BASE COUNT 3 a 8 c 7 g 4 t

ORIGIN

Query Match 100.0%; Score 22; DB 9; Length 22;

Best Local Similarity 100.0%; Pred. No. 15;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 3

LOCUS A48910 3402 bp DNA PAT 07-MAR-1997

DEFINITION Sequence 1 from Patent WO9605312.

ACCESSION A48910

VERSION A48910.1 GI:2302570

KEYWORDS

SOURCE Trypanosoma cruzi.

ORGANISM Trypanosoma cruzi.

REFERENCE 1 (bases 1 to 3402)

AUTHORS Paranhos-Baccala, G., Lesenechal, M. and Jolivet, M.

TITLE NOVEL TRYPA NOSOMA CRUZI ANTIGEN, AND GENE CODING THEREFOR, THEIR APPLICATION TO THE DETECTION OF CHAGAS' DISEASE

JOURNAL Patent: WO 9605312-A 1 22-FEB-1996;

COMMENT BIO MERIEUX (FR)

Other publication CA 2173957 960222

Other publication AU 3169195 960307

Other publication FR 2723589 960216.

FEATURES

source 1..3402

/organism="Trypanosoma cruzi"

/strain="G"

/db_xref="taxon:5693"

/dev_stage="EPIMASTIGOTE"

BASE COUNT 889 a 818 c 958 g 737 t

ORIGIN

Query Match 100.0%; Score 22; DB 9; Length 3402;

Best Local Similarity 100.0%; Pred. No. 3.5;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cagccgacgtagctgcgtcct 22

|||||

Db 1266 CAGCCGACGCTAGCTGCTCT 1287

RESULT 4
LOCUS AR047920 3402 bp DNA PAT 29-SEP-1999
DEFINITION Sequence 1 from patent US 5820864.
ACCESSION AR047920
VERSION AR047920.1 GI:5970263
KEYWORDS
SOURCE Unknown.
ORGANISM Unkown.
REFERENCE 1 (bases 1 to 3402)
AUTHORS Paranhos-Baccala,G., Lesenechal,M. and Jolivet,M.
TITLE Trypanosoma cruzi antigen, gene encoding therefor and methods of detecting and treating chagas disease
JOURNAL Patent: US 5820864-A 1 13-OCT-1998;
FEATURES
source location/Qualifiers
1..3402
BASE COUNT 888 a 821 c 956 g 737 t
ORIGIN

Query Match 100.0%; Score 22; DB 9; Length 3402;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cagccgacgtagctgcctcct 22
|||||
Db 1266 CAGCCGACGCTAGCTGCTCCT 1287

RESULT 5
LOCUS TC024190 3402 bp mRNA INV 04-AUG-1997
DEFINITION Trypanosoma cruzi Tc40 antigen (Tc40) mRNA, complete cds.
ACCESSION U24190
VERSION U24190.1 GI:790645
KEYWORDS
SOURCE Trypanosoma cruzi.
ORGANISM Trypanosoma cruzi.
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma; Schizotrypanum.
REFERENCE 1 (bases 1 to 3402)
AUTHORS Lesenechal,M., Duret,L., Cano,M.I., Mortara,R.A., Jolivet,M., Camargo,M.E., da Silveira,J.F. and Paranhos-Baccala,G.
TITLE Cloning and characterization of a gene encoding a novel immunodominant antigen of Trypanosoma cruzi
JOURNAL Mol. Biochem. Parasitol. 87 (2), 193-204 (1997)
97391123
2 (bases 1 to 3402)
AUTHORS Lesenechal,M., Franco Da Silveira,J., Mortara,R.A., Duret,L., Camargo,M.E., Jolivet,M. and Paranhos-Baccala,G.
TITLE Direct Submission
JOURNAL Submitted (06-APR-1995) Mylene Lesenechal, Biomerieux, 22 rue Saint Jean de Dieu, Lyon 69007, France
FEATURES
source location/Qualifiers
1..3402
/organism="Trypanosoma cruzi"
/strain="G"
/db_xref="taxon:5693"
/dev_stage="epimastigote"
1..35
/gene="Tc40"
/note="spliced leader"
1..3402
/gene="Tc40"
266..3013
/gene="Tc40"
/codon_start=1
/product="Tc40 antigen"
/protein_id="AAC47657.1"

/db_xref="GI:790646"
/translation="MFTVDFPNHAKPSNNEGRVSWSDATNEVPEARVLDASQFY
LATMKRRHRLVRVYKSNLTKGVRAKPIHAKVPYNVNSNVAASGAKKEEFVWYCT
DETETASNGKPDILARLTVKVFYFKLDQDPTIPCFSEFINASQRPDLVLLETQAAILD
SSSLIERFDESLEATLQKNCCTLRLLTPVSENSLCSVSGGMFTTETPTQVAICT
LRNRSPTSMACCEGEPAKALHLDATVEENSVLVAASGTGVOMLTGYAEPNLAK
FVIDGSIVAMSSRETFVAPDDKROLALVMHSPHNTCTHYMPCOVONGCFNPT
ADSCVILADMSNRLTFHLRSSRREDOPOKTSVATKPKGCVSGTDAASSHTPT
TSAAPASPAAPSPAPAKAAPAPAAASAPPHYSKILNVLNVLGINTVQRYVSTG
APATRTAVTSTTAPQRTSPYGHNGRPVYAGLVANSSASASASTPTAAKPTGEK
ASACETSSVAINATRPALHNASLPQAPDTGVLAAYVSGEYHSLERLESTVNT
SRVLEKLPDITIRDEQLNLGLAQTEIQOSPPFOTOPRDTSSAKSSVFETVTV
LIDLSLRNITKGVKGVNEMAIMLHDVEVHAIIGNLROTKNIRLSDLEKST
TOFPAOLTOVENVYKRELAEVIGSTGSLISVKNASIKELNISTSSGVDENMR
MREELCTRESVAKRKATMPDSSSHATSSQGRSAFETTLATRLSVRQQRQIGLE
YMLAQPSLLRLPLSTLTRENEVASSELLNEVETPDVNCVSLQDLTEAATAEKE
VVYGAVIDIISERQDIAONGALGSKLTPTAMRAFERARSETTSSFLQCKLNKELQ
S"

BASE COUNT 889 a 818 c 958 g 737 t
ORIGIN

Query Match 100.0%; Score 22; DB 96; Length 3402;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cagccgacgtagctgcctcct 22
|||||
Db 1266 CAGCCGACGCTAGCTGCTCCT 1287

RESULT 6
LOCUS AC021683 158566 bp DNA HNG 20-SEP-2000
DEFINITION Homo sapiens clone Rp11-13K12, WORKING DRAFT SEQUENCE, 14 unordered pieces.
ACCESSION AC021683
VERSION AC021683.3 GI:10198432
KEYWORDS HNG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 158566)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Unpublished
JOURNAL Homo sapiens, clone Rp11-13K12
2 (bases 1 to 158566)
AUTHORS Anderson,S., Baldwin,J., Bana,N., Beckerly,R., Beda,F., Boguslavsky,L., Bouhgalter,B., Brown,A., Burkett,G., Castle,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArrellano,K., Dewar,K., Domino,M., Doyle,M., Fenesfor,J., Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heatford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Kartas,A., Klein,J., Landers,T., Lehoczyk,U., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Margulis,N., McEwan,P., McGurk,A., McKernan,K., McPheters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K., Pierre,N., Pisanl,C., Pollara,V., Raymond,C., Riley,R., Rothman,D., Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Tjirell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wymann,D., Ye,W.J., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (19-JAN-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Sep 20, 2000 this sequence version replaced gi:7144663.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center

```

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information -----
Center project name: L3343
Center clone name: L3_K12
----- Summary Statistics -----
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.96071
Consensus quality: 152822 bases at least Q40
Consensus quality: 155339 bases at least Q30
Consensus quality: 156788 bases at least Q20
Insert size: 165000; agarose-fp
Insert size: 157266; sum-of-contigs
Quality coverage: 4.8 in Q20 bases; agarose-fp
Quality coverage: 5.0 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 1431: contig of 1431 bp in length
* 1432 1531: gap of 100 bp
* 1532 2879: contig of 1348 bp in length
* 2880 2979: gap of 100 bp
* 2980 6254: contig of 3275 bp in length
* 6255 6354: gap of 100 bp
* 6355 11675: contig of 5321 bp in length
* 11676 11775: gap of 100 bp
* 11776 16766: contig of 4991 bp in length
* 16767 16866: gap of 100 bp
* 16867 22571: contig of 5705 bp in length
* 22572 22671: gap of 100 bp
* 22672 30550: contig of 7879 bp in length
* 30551 30650: gap of 100 bp
* 30651 38928: contig of 8278 bp in length
* 38929 39028: gap of 100 bp
* 39029 47882: contig of 8854 bp in length
* 47883 47982: gap of 100 bp
* 47983 60422: contig of 12440 bp in length
* 60423 60522: gap of 100 bp
* 60523 75477: contig of 14955 bp in length
* 75478 75577: gap of 100 bp
* 75578 92875: contig of 17298 bp in length
* 92876 92975: gap of 100 bp
* 92976 130172: contig of 37197 bp in length
* 130173 130272: gap of 100 bp
* 130273 158566: contig of 28294 bp in length.
-----
FEATURES
source
1..158566
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Rp11-13K12"
/clone.lib="RPC1-11 Human Male BAC"
1..1431
/note="assembly-fragment"
clone_end:SP6
vector_side:left"
misc_feature
1532..2879
/note="assembly-fragment"
2980..6254
/note="assembly-fragment"
6355..11675
/note="assembly-fragment"
11776..16766
/note="assembly-fragment"
16867..22571
misc_feature

```

```

misc_feature /note="assembly-fragment"
22672..30550
/note="assembly-fragment"
misc_feature 30651..38928
/note="assembly-fragment"
misc_feature 39029..47882
/note="assembly-fragment"
misc_feature 47983..60422
/note="assembly-fragment"
misc_feature 60523..75477
/note="assembly-fragment"
misc_feature 75578..92875
/note="assembly-fragment"
misc_feature 92976..130172
/note="assembly-fragment"
misc_feature 130273..158566
/note="assembly-fragment"
clone_end:77
vector_side:right"
BASE COUNT 40069 a 38517 c 37765 g 40914 t 1301 others
ORIGIN
Query Match 80.9%; Score 17.8; DB 66; Length 158566;
Best Local Similarity 90.5%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 caacgcagctactcgcgc 21
Db 61377 CACCTACGACGACCTCGTCC 61357
-----
RESULT 7
AC005439
LOCUS
DEFINITION
AC005439
AC005439
AC005439.1 GI:3548783
HTG.
Drosophila melanogaster (Subclones in sac from P1 clone DS01989
(D353)) DNA.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 29485)
REFERENCE
Celiker,S.E., George,R.A., Galle,R.F., Hoskins,R.A.,
Svirskas,R.R., Harris,N.L., Agbayani,A., Arcaina,T.T., Baxter,E.,
Blazej,R.G., Chavez,C., Chew,M., Doyle,C.M., Farfan,D.E.,
Flanagan,J., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,S.H., Lee,B., Lomotan,M.A., Mak,J., Mazda,P., Moshrefi,A.R.,
Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,E.,
Snir,E., Twomey,B., Wan,K.H., Zhang,R., Zieran,L.L. and Rubin,G.M.
Snir,E., Twomey,B., Wan,K.H., Zhang,R., Zieran,L.L. and Rubin,G.M.
Sequencing of Drosophila chromosome 2R, region 46F5-46F7
Unpublished (1997)
2 (bases 1 to 29485)
REFERENCE
Celiker,S.E., George,R.A., Galle,R., Svirskas,R.R., Hoskins,R.A.,
Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Chavez,C.,
Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A.,
Hummasti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B.,
Lomotan,M.A., Mak,J., Mazda,P., MOK,M.S., Moshrefi,A.R.,
Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,E.,
Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R.,
Zieran,L.L. and Kimmel,B.E.
Direct Submission
Submitted (14-AUG-1998) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Sep 5, 1998 this sequence version replaced gl:3419743.
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
TITLE
JOURNAL
COMMENT

```


For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (<http://fruitfly.berkeley.edu/sequence/>) or send email to bdg@fruitfly.berkeley.edu.
Library Location: 69-21.

FEATURES
source
Location/Qualifiers
1..29485

/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="2R"
/map="46F5-46F7"
/clone="PI D501989 (D353)"

BASE COUNT 8297 a 6045 c 6384 g 8759 t
ORIGIN

Query Match 78.2%; Score 17.2; DB 4; Length 29485;
Best Local Similarity 86.4%; Pred. No. 3.4e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 cagcgcagcgtacgtcgtcct 22
||||| ||||| ||||| ||||| |||||
Db 21681 CAGCTGACCGAGCTGCTCCT 21702

RESULT 8
LOCUS AC017302/c
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered pieces.
ACCESSION AC017302
VERSION AC017302.1 GI:6553684
KEYWORDS HTG: HTGS_PHASE2.
SOURCE fruit fly.

ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 33077)
Adams, M. and Venter, J.C.

AUTHORS
TITLE Direct Submission
JOURNAL Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA
COMMENT This sequence was identified as CDM:10210213 by the submitter. For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
source
Location/Qualifiers
1..33077

/organism="Drosophila melanogaster"
/db_xref="taxon:7227"

BASE COUNT 9543 a 7246 c 7112 g 9176 t
ORIGIN

Query Match 78.2%; Score 17.2; DB 64; Length 33077;
Best Local Similarity 86.4%; Pred. No. 3.3e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 cagcgcagcgtacgtcgtcct 22
||||| ||||| ||||| ||||| |||||
Db 21449 CAGCTGACCGAGCTGCTCCT 21428

RESULT 9
LOCUS HSJ858B16
DEFINITION Human DNA sequence from clone RPS-858B16 on chromosome 22 contains the KIAA0542 gene and the P1SD gene for phosphatidylserine decarboxylase (PSSC, EC 4.1.1.65). Contains ESTs, GSSs and a putative Cpg island, complete sequence.

ACCESSION AL096768
VERSION AL096768.7 GI:5596770
KEYWORDS HTG: CPG island; KIAA0542; Phosphatidylserine Decarboxylase
Proenzyme; P1SD; PSSC.
human.

SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 36676)
Barlow, K.
Direct Submission
Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
On Jul 27, 1999 this sequence version replaced gi:5579004.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr22>

Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; SW, SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep RPS-858B16 is from the library RPS-5 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/> VECTOR: pCFPAC2

IMPORTANT: This sequence is not the entire insert of clone RPS-858B16. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

The true left end of clone RPS-858B16 is at 1 in this sequence. The true left end of clone RPS-694E4 is at 36572 in this sequence.

FEATURES
source
Location/Qualifiers
1..36676

/organism="Homo sapiens"
/db_xref="taxon:9606"

/chromosome="22"
/clone_11b="RPS-5"
/clone="RPS-858B16"
join(<273..454,1413..1488,1978..2073,2538..2614,3997..4099,5606..5702,8812..8972,9435..9509,10811..10885,11023..11199,11271..11524,12717..12948,14414..14546,14669,14747,15784..15895,15983..16218)
/gene="dJ858B16.1"

/note="match: cDNAs: Em:AB011114"
/evidence="not-experimental"
/product="dJ858B16.1 (KIAA0542 (isoform 2))"
join(<273..454,1413..1488,1978..2073,2538..2614,3997..4099,5606..5702,8812..8972,9435..9509,10811..10885,11023..11199,11271..11524,12717..12948,14414..14546,14669..14747,15784..15895,15983..16217)
/gene="dJ858B16.1"

/note="match: ESTs: Em:Z78408 Em:R53362 Em:AI806913
Em:AA521233 Em:R53964 Em:AA158770 Em:AI858196 Em:AA594535
Em:AA836285 Em:AA974403 Em:AI425026 Em:AI809747
Em:AI978557 Em:AI141022 Em:AA159572 Em:AA535345
Em:AA582394 Em:RS1102 Em:AA829254 Em:R54651 Em:AA971234
Em:AA846853 Em:N56896 Em:W06066 Em:AA118158 Em:AV030271
Em:AV116402 Em:AI024124 Em:AV122327 Em:N63934 Em:AI191287

CDS	protein_id="AACT0187.1" /translation="MFKSYAGNEPESNLFHNEAYPPREDALRLVSEVAAAAPASAOQL /db_xref="GI:3822236" /codon_start=1 /product="protein kinase" /protein_id="AAC70188.1" /db_xref="GI:3822237" /translation="MDALIGDFADPFKECSARTALHLVNGKCKSVWKKHCPQKSPF IAKRIEKKHNAIEPVHLMKFNKYFLRLFLYSLHSREHLVMDVYIPDDLEFLDMOTE PMLREBELSLIAYOLIDALQALHKHNVANDLVNLYLRQOIVYCYGLCKEIS PSTFGATVLYSEPERINKHAAVHPDMAVGVLLIETSGKHFFKIDODESLDEVLH KROIADVPEPDNPFLEEFICFLGYCYDRAHSYEVIOKNTYWKSIWHMKR" complement(4501..7590) /note="109196 Da; Ldof-4" /codon_start=1 /product="mucin-like protein" /protein_id="AAC70189.1" /db_xref="GI:3822238" /translation="MCKNCAFYIITKNLKAISEVDPDSDGSENERLILVNRGCGTD ADPGSPKSDKSSDSDROSKSKSDSSSDSEDEPERDPOMLIIPVEDKILNLR KILKLINERAEAIKMLRDOQAPPEVPPROKSDSSSSSSSSGSDSDSDDECT KNVSKNFTEIYSPREFEAKNVSEDLFEATGESEIAEPPEAPADAEPAEPPEAE PADAPAPDESALEEPVSPSSCLISPGPMOFEVDEPATSPGEHKPOPLSNKA STVSDKSSPSADKHSRDKSPKSGCGRPKSSVSGKPKPEAKSKSCDKPQPG VSTSGYDKRKKSSADKPSDKAKSSSTVKKSSADPTPSDKAKSSSTVKKSSSDKSD KAKSSSSDOKPCDKAKSSSTVKKSSPDKPCDKAKSSSTVKKSSSDKSPDKLSDKPS DKAKSSSTVKKSSPDKSPDKAKSSSTVKKSSPDKSPDKAKSSSTVKKSSPDKSPD KAKSSSTVKKSSAPDKPKHSSDKSDATVYKSSSAVPDKSPDKSSDGSFRSP PESEDRDQSLAOTSPPFLAALARDMDERLNLDAVAATIGVAPPEYPAPEPA PYDPTLNAEEFAPAPPEPAPEPAPEPAPEPAPEPAPEPAPEPAPEPAPEPAPE TSRTSEPALGVEPALPEVEPALPEVEPALGEPALGEPALGEPALGEPALGEPALGEP AVERPAQAPDLTSDAVESTFGKSTVASELAARPAELAAPELPEHESEPAQ PASKPAPEPAHLTPDRLTLPPLVSKSLMSKSDDEVEVEDIFVNEDEDEVED IYIKNEMLCLVGVGDEYEHKASSSTPSAAVEVDEVEIMSEFPMENHAKPCVFMV KKKSDMLNVRKRPYSROOPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP HLPQSPFVPPQPFPRYVATPALPDHCAPEAVPHQEVLDLITLCGPDPTTPON" 7722..8291 /note="21763 Da; Ldof-5" /codon_start=1 /product="Ldof-5 peptide" /protein_id="AAC70190.1" /db_xref="GI:3822239" /translation="MKTIVPHDRHILFDPAVDHLDRVGVVRYIIHAYFLGHVSVFI QLEHAFOVILTLHDHTVYVFLGLESFLGYDLFVFLVYTKLIDLVLHVOAETLQOI VLOTLLHSHVTLVHLVGHVPAVESYLKEVDPYKLEAVRFONGHHVKLLAVLQ NVYVLYVLIYVEADLPLGLIARSVY" complement(8282..8524) /note="8960 Da; Ldof-6" /codon_start=1 /product="Ldof-6 peptide" /protein_id="AAC70191.1" /db_xref="GI:3822240" /translation="MSEPPGRGRANGAQTADAAARLLIARLLIVSIGRLIASRT REGIDGSTRIRGKLSIRVYDRAIVLQFLTCY" 8619..9194 /note="22146 Da; Ld-922; Ldof-7" /codon_start=1 /product="Ldof-7 peptide" /protein_id="AAC70192.1" /db_xref="GI:3822241" /translation="MSLRTFNHVAQVYLLTRDGRFSVTKTVNCLFMERKKNRDVY GQYVRKSDSKTKVCKGYVLCKYQYMCVHDKLVTGTRDFLKRFNVVPPPGVAVYQ
CDS	protein_id="AACT0187.1" /translation="MFKSYAGNEPESNLFHNEAYPPREDALRLVSEVAAAAPASAOQL /db_xref="GI:3822236" /codon_start=1 /product="protein kinase" /protein_id="AAC70188.1" /db_xref="GI:3822237" /translation="MDALIGDFADPFKECSARTALHLVNGKCKSVWKKHCPQKSPF IAKRIEKKHNAIEPVHLMKFNKYFLRLFLYSLHSREHLVMDVYIPDDLEFLDMOTE PMLREBELSLIAYOLIDALQALHKHNVANDLVNLYLRQOIVYCYGLCKEIS PSTFGATVLYSEPERINKHAAVHPDMAVGVLLIETSGKHFFKIDODESLDEVLH KROIADVPEPDNPFLEEFICFLGYCYDRAHSYEVIOKNTYWKSIWHMKR" complement(4501..7590) /note="109196 Da; Ldof-4" /codon_start=1 /product="mucin-like protein" /protein_id="AAC70189.1" /db_xref="GI:3822238" /translation="MCKNCAFYIITKNLKAISEVDPDSDGSENERLILVNRGCGTD ADPGSPKSDKSSDSDROSKSKSDSSSDSEDEPERDPOMLIIPVEDKILNLR KILKLINERAEAIKMLRDOQAPPEVPPROKSDSSSSSSSSGSDSDSDDECT KNVSKNFTEIYSPREFEAKNVSEDLFEATGESEIAEPPEAPADAEPAEPPEAE PADAPAPDESALEEPVSPSSCLISPGPMOFEVDEPATSPGEHKPOPLSNKA STVSDKSSPSADKHSRDKSPKSGCGRPKSSVSGKPKPEAKSKSCDKPQPG VSTSGYDKRKKSSADKPSDKAKSSSTVKKSSADPTPSDKAKSSSTVKKSSSDKSD KAKSSSSDOKPCDKAKSSSTVKKSSPDKPCDKAKSSSTVKKSSSDKSPDKLSDKPS DKAKSSSTVKKSSPDKSPDKAKSSSTVKKSSPDKSPDKAKSSSTVKKSSPDKSPD KAKSSSTVKKSSAPDKPKHSSDKSDATVYKSSSAVPDKSPDKSSDGSFRSP PESEDRDQSLAOTSPPFLAALARDMDERLNLDAVAATIGVAPPEYPAPEPA PYDPTLNAEEFAPAPPEPAPEPAPEPAPEPAPEPAPEPAPEPAPEPAPEPAPE TSRTSEPALGVEPALPEVEPALPEVEPALGEPALGEPALGEPALGEPALGEPALGEP AVERPAQAPDLTSDAVESTFGKSTVASELAARPAELAAPELPEHESEPAQ PASKPAPEPAHLTPDRLTLPPLVSKSLMSKSDDEVEVEDIFVNEDEDEVED IYIKNEMLCLVGVGDEYEHKASSSTPSAAVEVDEVEIMSEFPMENHAKPCVFMV KKKSDMLNVRKRPYSROOPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP HLPQSPFVPPQPFPRYVATPALPDHCAPEAVPHQEVLDLITLCGPDPTTPON" 7722..8291 /note="21763 Da; Ldof-5" /codon_start=1 /product="Ldof-5 peptide" /protein_id="AAC70190.1" /db_xref="GI:3822239" /translation="MKTIVPHDRHILFDPAVDHLDRVGVVRYIIHAYFLGHVSVFI QLEHAFOVILTLHDHTVYVFLGLESFLGYDLFVFLVYTKLIDLVLHVOAETLQOI VLOTLLHSHVTLVHLVGHVPAVESYLKEVDPYKLEAVRFONGHHVKLLAVLQ NVYVLYVLIYVEADLPLGLIARSVY" complement(8282..8524) /note="8960 Da; Ldof-6" /codon_start=1 /product="Ldof-6 peptide" /protein_id="AAC70191.1" /db_xref="GI:3822240" /translation="MSEPPGRGRANGAQTADAAARLLIARLLIVSIGRLIASRT REGIDGSTRIRGKLSIRVYDRAIVLQFLTCY" 8619..9194 /note="22146 Da; Ld-922; Ldof-7" /codon_start=1 /product="Ldof-7 peptide" /protein_id="AAC70192.1" /db_xref="GI:3822241" /translation="MSLRTFNHVAQVYLLTRDGRFSVTKTVNCLFMERKKNRDVY GQYVRKSDSKTKVCKGYVLCKYQYMCVHDKLVTGTRDFLKRFNVVPPPGVAVYQ
CDS	protein_id="AACT0187.1" /translation="MFKSYAGNEPESNLFHNEAYPPREDALRLVSEVAAAAPASAOQL /db_xref="GI:3822236" /codon_start=1 /product="protein kinase" /protein_id="AAC70188.1" /db_xref="GI:3822237" /translation="MDALIGDFADPFKECSARTALHLVNGKCKSVWKKHCPQKSPF IAKRIEKKHNAIEPVHLMKFNKYFLRLFLYSLHSREHLVMDVYIPDDLEFLDMOTE PMLREBELSLIAYOLIDALQALHKHNVANDLVNLYLRQOIVYCYGLCKEIS PSTFGATVLYSEPERINKHAAVHPDMAVGVLLIETSGKHFFKIDODESLDEVLH KROIADVPEPDNPFLEEFICFLGYCYDRAHSYEVIOKNTYWKSIWHMKR" complement(4501..7590) /note="109196 Da; Ldof-4" /codon_start=1 /product="mucin-like protein" /protein_id="AAC70189.1" /db_xref="GI:3822238" /translation="MCKNCAFYIITKNLKAISEVDPDSDGSENERLILVNRGCGTD ADPGSPKSDKSSDSDROSKSKSDSSSDSEDEPERDPOMLIIPVEDKILNLR KILKLINERAEAIKMLRDOQAPPEVPPROKSDSSSSSSSSGSDSDSDDECT KNVSKNFTEIYSPREFEAKNVSEDLFEATGESEIAEPPEAPADAEPAEPPEAE PADAPAPDESALEEPVSPSSCLISPGPMOFEVDEPATSPGEHKPOPLSNKA STVSDKSSPSADKHSRDKSPKSGCGRPKSSVSGKPKPEAKSKSCDKPQPG VSTSGYDKRKKSSADKPSDKAKSSSTVKKSSADPTPSDKAKSSSTVKKSSSDKSD KAKSSSSDOKPCDKAKSSSTVKKSSPDKPCDKAKSSSTVKKSSSDKSPDKLSDKPS DKAKSSSTVKKSSPDKSPDKAKSSSTVKKSSPDKSPDKAKSSSTVKKSSPDKSPD KAKSSSTVKKSSAPDKPKHSSDKSDATVYKSSSAVPDKSPDKSSDGSFRSP PESEDRDQSLAOTSPPFLAALARDMDERLNLDAVAATIGVAPPEYPAPEPA PYDPTLNAEEFAPAPPEPAPEPAPEPAPEPAPEPAPEPAPEPAPEPAPEPAPE TSRTSEPALGVEPALPEVEPALPEVEPALGEPALGEPALGEPALGEPALGEPALGEP AVERPAQAPDLTSDAVESTFGKSTVASELAARPAELAAPELPEHESEPAQ PASKPAPEPAHLTPDRLTLPPLVSKSLMSKSDDEVEVEDIFVNEDEDEVED IYIKNEMLCLVGVGDEYEHKASSSTPSAAVEVDEVEIMSEFPMENHAKPCVFMV KKKSDMLNVRKRPYSROOPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP HLPQSPFVPPQPFPRYVATPALPDHCAPEAVPHQEVLDLITLCGPDPTTPON" 7722..8291 /note="21763 Da; Ldof-5" /codon_start=1 /product="Ldof-5 peptide" /protein_id="AAC70190.1" /db_xref="GI:3822239" /translation="MKTIVPHDRHILFDPAVDHLDRVGVVRYIIHAYFLGHVSVFI QLEHAFOVILTLHDHTVYVFLGLESFLGYDLFVFLVYTKLIDLVLHVOAETLQOI VLOTLLHSHVTLVHLVGHVPAVESYLKEVDPYKLEAVRFONGHHVKLLAVLQ NVYVLYVLIYVEADLPLGLIARSVY" complement(8282..8524) /note="8960 Da; Ldof-6" /codon_start=1 /product="Ldof-6 peptide" /protein_id="AAC70191.1" /db_xref="GI:3822240" /translation="MSEPPGRGRANGAQTADAAARLLIARLLIVSIGRLIASRT REGIDGSTRIRGKLSIRVYDRAIVLQFLTCY" 8619..9194 /note="22146 Da; Ld-922; Ldof-7" /codon_start=

/codon_start=1
/protein_id="BAA8533.1"
/db_xref="GI:6630687"
/translation="MAPPAMAEAKVSCPLAREVGRRAAAGGCGRNFTVSPISPH
ERETAAATTAAGVAGDVGTGGAHSEKAGHRSNTTLVCTWTWYSEVSGSGENL
CRTEGRLNTLVSSAGSVVMKGDAETVETQEWENLDTTEEMCRIMWGNONSQCE
RCIKKRRITTAATATAKNKODILTLA"
Join(61440..63104,63219..63302,63477..63581,63692..63800,
64338..64489,64592..64738)
/note="Similar to Arabidopsis thaliana DNA chromosome 4,
BAC clone T5L19 (AU049481)"
/codon_start=1
/protein_id="BAA8534.1"
/db_xref="GI:6630688"
/translation="MDSRSPKIPRRSPDSKDKSDNRKKEADKNDMSRAGSEETDC
KEEMCSNKRKGLTNGEIVYDNRSDSCHEPTLHLYLRDRQKSVETIKILIDGVAK
SDYAKQKIDLSERRNGTGNRSVDYLRDKLDSGRDRMSDRTREBEGSKDYVRNQ
WDSKANDSEKNAHERLDGSGFHGRAGYRDSRGSSEIRSSYTGGRYDSDSIE
IRPNMNLDFREGSVSGRRYDGAHNDATPGTNGDSANEADSGSTSTISOPQHG
PKGDRSSRGGRNSRDSQVGGTLPMPPPGLPLPMPQHIGHIPSQGPLP
GVFPVPPGLVWAGRVNMLSVPMNPIPVAAEHFASMGACGPHNHLNLI
GSGTGAPTVSGSLFHQLOSREMAHDKPPACGWTPHNSGPTKAPRGQNDYS
ONFVDIGMRPONTREDLTSVADYPKLELQKDEIVANSAPMYKCDLRQY
LSEFFGTGFEDVILVDPMEVYHRAPIGTIDIEYNAEIMNLKLEALDISEFVL
WVDGGLLEGROQLKMGFRCEVCWTKNKATPSLRHSHSTILQSHKCLMG
IKGTVRSTGHIYHANIDTDIIAEPTDGTSTKPEDMYRIIEHFALGRLELGE
DNIIRPGLTGLGKLSYNNKRAYIKNFADKQKXWGGCGGNPPENHLYVTPE
IEGRKSPPOKN"
Join(5836..65662,66040..66217,67638..67719,68289..68465,
68556..68723,68848..69039)
/note="ESTs D39011(R0609), AU032023(R3215) correspond to a
region of the predicted gene.
Similar to 26S proteasome, non-ATPase subunit. (Y13071)"
/codon_start=1
/protein_id="BAA8535.1"
/db_xref="GI:6630689"
/translation="MERLORIFGASGNGCPPTDPLIDSEQYVYSSALLMLKHKGR
AGVPMVGMILGEFVDYTVRVADVAFAMQSGTGVSEVADVFQNMIDMKORR
PEVVGVMHSHPGFCMLSGVDINTOOSFALNPAVAVYIDPLOSRYKGVYIDAPFL
INQTMILCOEPROTSSNGHLKPSIQALIHNLNRYISIAINRKNLEEMKMLNL
HKRWTDGLIKSFDSHTKNEOTVOEMLAIKYNAAVEDELEPEKLIANVQKQ
DARKHLEHVSNNMSSNIVQTLGTMIDTVF"
71085..72281
/note="Similar to serpin - wheat. (565782)"
/codon_start=1
/protein_id="BAA8536.1"
/db_xref="GI:6630690"
/translation="MAPPAMAEAKVSCPLAREVGRRAAAGGCGRNFTVSPISPH
AALAVDAGRGCTORELLGLFGLSPSLAEHRSPTRLVRLHLPTSPACGVWR
GRALPEPDAASRYAAVAPADPAPOAERNAFVSDTEGIRVLPNSVD
SSYVVLVANAHRKGTWSLDPHPSATPHAFPHLLDGAVALPMTIEPERRVAAF
GFTALKLPYKNNVGGGCGDVPRAAFTMLLLPDGALKLADLIDMAVITPERIKH
TPAAEAPVRLVMAKFSFKFEAKSDMRKLVRAFGDFGSMVTGGGLTALEY
HQTIEVDELGTVAASATAVVMQKSSLPDPVADVADRPFLFVVELELCAVLGIV
VNLAE"
complement(Join(72911..73060,73798..74024,74469..74535))
/note="ESTs AU082078(C53435), C28061(C53846) correspond to
a region of the predicted gene.
Similar to Pgp24. (AF049928)"
/codon_start=1
/protein_id="BAA8537.1"
/db_xref="GI:6630691"
/translation="MIRHACYSLTCDLSLSEKIKGTCATPCPLFGNAPLGS
TLAGSTTHOMLGLTSLCCVFTGGVLAIVPGSAVACIAGTGRALRAKATNLEAAC
GDLTFLFCHLCAICQERIRERTDSSGSAPTVTPAIQTMDEP"
complement(Join(81876..81979,82580..82833,83032..83196,
83667..83799,83939..84043,84886..85016,85649..85742))
/note="hypothetical protein"
/codon_start=1

Query Match 78.28; Score 17.2; DB 13; Length 168655;
Best Local Similarity 86.4%; Pred. No. 2e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 cagcgcagctagctgcgctct 22
11 ||||| ||||| |||||
Db 11325 CAACGACGACGCTGCTCCT 11304

RESULT 14

AC007417

LOCUS

DEFINITION

AC007417

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNALS

AUTHORS

TITLE

JOURNALS

AUTHORS

TITLE

JOURNALS

AUTHORS

TITLE

JOURNALS

AUTHORS

TITLE

JOURNALS

AUTHORS

TITLE

JOURNALS

AUTHORS

TITLE

JOURNALS

AUTHORS

TITLE

JOURNALS

AUTHORS

TITLE

JOURNALS

AUTHORS

TITLE

JOURNALS

AUTHORS

TITLE

JOURNALS

AUTHORS

TITLE

JOURNALS

AUTHORS

TITLE

JOURNALS

AUTHORS

TITLE

JOURNALS

AUTHORS

TITLE

JOURNALS

AUTHORS

BASE COUNT 49836 a 38215 c 38034 g 49597 t

AC007417 175682 bp DNA INV 02-MAR-2001
Drosophila melanogaster, chromosome 2R, region 47A-47B, BAC clone
BACR48F07, complete sequence.
AC007417
AC007417.4 GI:13184056
HTG.
fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscophora; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 175682)
Celisner,S.E., Adams,M.D., Krommiller,B., Tyler,D., Wan,K.H.,
Holt,R.A., Evans,C.A., Gocayne,D., Amaratunga,P.G., Brannon,R.C.,
Rogers,Y., An,H., Baldwin,D., Bonzon,J., Beeson,K.Y., Busan,D.A.,
Carlson,J.W., Center,A., Champagne,M., Davenport,L.B., Dietz,S.M.,
Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
Ferreira,S., Fritze,E., Galle,R.F., Gar,N.S., George,R.A.,
Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
Idegawa,C., Jallil,M., Kruse,D., Li,P., Mallet,B., Moshrefi,A.,
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
Paclet,B., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
Phouanavong,S., Piltman,G.S., Puri,V., Richards,S., Scheeler,F.,
Stapleton,M., Strong,R., Svirskaas,R., Tector,C., Williams,S.M.,
Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
Sequencing of Drosophila chromosome 2R, region 47A-47B
Unpublished
2 (bases 1 to 175682)
Celisner,S.E., Abghayani,A., Arcana,T.T., Baxter,E., Blazet,R.G.,
Baltenhoff,C., Champagne,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Katra,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomutan,M.A., Mazda,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Paclet,B., Park,S.,
Pfeiffer,B., Poon,L., Sequelira,A., Sethi,H., Snit,E.,
Svirskaas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zietan,L.L. and
Rubin,G.M.
Direct Submission
Submitted (28-APR-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Mar 2, 2001 this sequence version replaced gi:5670592.
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
This sequence was assembled using end sequences from a whole genome
shotgun and from subclones of this BAC and its neighboring clones.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive web site (<http://www.fruitfly.org/sequence/>) or send email
to bdg@fruitfly.berkeley.edu.
Location/Qualifiers
1. 175682
/organism="Drosophila melanogaster"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/chromosome="2R"
/map="47A-47B"
/clone="BACR48F07 (D625)"
/clone.lib="RPCI-98 (Rockwell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
pBACe3.6)"


```
11235..11542,11596..>12210)
/gene="CG11866"
/product="Ct37014"
/db_xref="FLYBASE:FBan0011866"
/db_xref="FLYBASE:FBgn0033486"
<8915..>12210
/gene="CG11866"
/map="46E4-46E6"
/db_xref="FLYBASE:FBan0011866"
/db_xref="FLYBASE:FBgn0033486"
join(9286..9726,10663..10911,10976..11174,11235..11542,
11596..12210)
/gene="CG11866"
/note="CG11866 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0011866"
/db_xref="FLYBASE:FBgn0033486"
/protein_id="AAF58839.1"
/db_xref="GI:7303791"
/translation="MVSTRQMSGGTSSAEGVSPFGAGSGSASGTGMAAGSASVSG
SGTVGPEPSTANRSSVLRRSTSYGNOAHSSRYVASGSLAESAVTRLHSYTMQ
MORQRPPTATTTTTSAPSLYDLPNELIEKILSYDYKKVSNRLVSHRNDICMAML
MTAFYKQITTTTSRFQAIKASMPRESHRNHPLACECDIIFTCYMRSLSLQMSGKH
IERGHCCFPFGALIDVOAILNYISITPRQRPYRTDELFLSTMAEYFKDRIEAT
LGLAYFNKDFYLPPTTKRPTLAISSLEDASNSPQNHVLRKGIKIKQGMKY
NNQSLVRLRELRTCKRKAASQSLAEOSSLISEQOKOTLEVANRLDENDXKNEEMR
KPSLLOELKCKTELQFWRKSPATPAVSCNOKAPVPPEPQALYNGVDEPH
FTIINEDVDESIVSVELKEPSPDESTTAKLAVSTAKNLKRMPSSESQNAIAS
TSAAEVQSOQSPMATGNAGAAKAGVSPKLFYGNHORSIVSPVSQNLGVAATT
DNVAPGEALIEPEEAKKARVOKANRYVTPGKRK"
complement(join(<12115..12937,12999..13642,13694..13708))
/gene="CG17753"
/product="Ct11457"
/db_xref="FLYBASE:FBan0017753"
/db_xref="FLYBASE:FBgn0033487"
complement(<12115..>13708)
/gene="CG17753"
/map="46E4-46E6"
/db_xref="FLYBASE:FBan0017753"
/db_xref="FLYBASE:FBgn0033487"
complement(join(12802..12937,12999..13642,13694..13708))
/gene="CG17753"
/note="CG17753 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0017753"
/db_xref="FLYBASE:FBgn0033487"
/protein_id="AAF58838.1"
/db_xref="GI:7303790"
/translation="MSSIKIEFAVQMRGDESYAGALRSALDVGQVEIDTQGRVYI
OTGRPMSIODKIEATGVRVAVLSGFGGSAVALINTGSVNDKPTIOGVVRFPTITAD
KRGVYVDGVGLSPGLCHLTHESDPTAGCSSYGEHNPQSPHGSAPAAEERH
AGDLGNIRADENGRATFRFVDPVLEYWDIIGRAVVLJANADDLGKGNQDSTLDGNSG
ERTACGIIARSAGILNFKRICACDGTWLDERNKPLACKERSQKL"
join(<18636..18761,19765..19876,19968..20213,22806..22878,
22958..23344,23489..23633,23934..24407,25026..>25202)
/gene="CG11867"
/product="Ct37016"
/db_xref="FLYBASE:FBan0011867"
/db_xref="FLYBASE:FBgn0033488"
<18636..>25202
/gene="CG11867"
/map="46E4-46E8"
/db_xref="FLYBASE:FBan0011867"
/db_xref="FLYBASE:FBgn0033488"
join(18636..18761,19765..19876,19968..20213,22806..22878,
22958..23344,23489..23633,23934..24407,25026..25202)
Query Match 78.2% Score 17.2; DB 5; Length 261846;
Best Local Similarity 86.4%; Pred. NO. 1.8e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
1 cagcgacggtagctgcctct 22
||||| ||| | ||||| |||||
```

Db 125602 CAGCTGACCGAGCTGCCTCT 125623

Search completed: September 21, 2001, 00:35:41
Job time: 29596 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 21, 2001, 02:42:38 : Search time 330.77 Seconds
(without alignments)
41.763 Million cell updates/sec

Title: US-09-138-735-9

Perfect score: 22

Sequence: 1 cagccgacgtagctgctct 22

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N.Geneseq_0601:*

1:	/cgnl_9/gcgdata/geneseq/geneseq/NA1980.DAT:*
2:	/cgnl_9/gcgdata/geneseq/geneseq/NA1981.DAT:*
3:	/cgnl_9/gcgdata/geneseq/geneseq/NA1982.DAT:*
4:	/cgnl_9/gcgdata/geneseq/geneseq/NA1983.DAT:*
5:	/cgnl_9/gcgdata/geneseq/geneseq/NA1984.DAT:*
6:	/cgnl_9/gcgdata/geneseq/geneseq/NA1985.DAT:*
7:	/cgnl_9/gcgdata/geneseq/geneseq/NA1986.DAT:*
8:	/cgnl_9/gcgdata/geneseq/geneseq/NA1987.DAT:*
9:	/cgnl_9/gcgdata/geneseq/geneseq/NA1988.DAT:*
10:	/cgnl_9/gcgdata/geneseq/geneseq/NA1989.DAT:*
11:	/cgnl_9/gcgdata/geneseq/geneseq/NA1990.DAT:*
12:	/cgnl_9/gcgdata/geneseq/geneseq/NA1991.DAT:*
13:	/cgnl_9/gcgdata/geneseq/geneseq/NA1992.DAT:*
14:	/cgnl_9/gcgdata/geneseq/geneseq/NA1993.DAT:*
15:	/cgnl_9/gcgdata/geneseq/geneseq/NA1994.DAT:*
16:	/cgnl_9/gcgdata/geneseq/geneseq/NA1995.DAT:*
17:	/cgnl_9/gcgdata/geneseq/geneseq/NA1996.DAT:*
18:	/cgnl_9/gcgdata/geneseq/geneseq/NA1997.DAT:*
19:	/cgnl_9/gcgdata/geneseq/geneseq/NA1998.DAT:*
20:	/cgnl_9/gcgdata/geneseq/geneseq/NA1999.DAT:*
21:	/cgnl_9/gcgdata/geneseq/geneseq/NA2000.DAT:*
22:	/cgnl_9/gcgdata/geneseq/geneseq/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	22	100.0	22	17	AAT27313
2	22	100.0	22	20	AAx84095
3	22	100.0	3402	17	AAT27310
4	22	100.0	3402	20	AAx84092
5	16.2	73.6	123	21	AAC24763
6	16.2	73.6	10475	9	AAAB0616
7	16.2	73.6	10475	9	AAAB0643
8	15.8	71.8	282	18	AAT91520
9	15.8	71.8	282	18	AAT91454
10	15.8	71.8	282	19	AAV64502
11	15.8	71.8	282	19	AAV44394

12	15.8	71.8	282	20	AAZ19304	M. tuberculosis an
13	15.8	71.8	282	20	AAZ19092	M. tuberculosis re
14	15.8	71.8	404	22	AAE68508	Human lung tumour
15	15.6	70.9	1661	21	AAK37040	Human PRO1772 (UNQ
16	15.6	70.9	1661	22	AAE54245	DNA encoding prote
17	15.6	70.9	1661	22	AAE76843	Human ORFX ORF2398
18	15.6	70.9	14327	22	AAE54724	Nucleotide sequenc
19	15.4	70.0	1262	21	AAE39276	Archidopsis thalia
20	15.2	69.1	136	18	AAE49467	AAV sequence joine
21	15.2	69.1	136	21	AAE09815	Human secreted pro
22	15.2	69.1	355	18	AAE91514	Mycobacterium tube
23	15.2	69.1	355	18	AAE91450	Mycobacterium tube
24	15.2	69.1	355	19	AAE64497	M. tuberculosis im
25	15.2	69.1	355	19	AAE44389	Mycobacterium tube
26	15.2	69.1	355	20	AAE19289	M. tuberculosis an
27	15.2	69.1	355	20	AAE19087	M. tuberculosis re
28	15.2	69.1	790	19	AAE24111	Human tumour prote
29	15.2	69.1	883	21	AAE56131	Eucalyptus grandis
30	15.2	69.1	1146	16	AAE08690	Creatine-kinase su
31	15.2	69.1	1437	21	AAE18037	Lung cancer associ
32	15.2	69.1	1539	19	AAE64478	M. tuberculosis im
33	15.2	69.1	1539	19	AAE44370	Mycobacterium tube
34	15.2	69.1	1539	20	AAE19280	M. tuberculosis an
35	15.2	69.1	1539	20	AAE19068	M. tuberculosis re
36	15.2	69.1	1568	14	AAE051620	Human creatine kin
37	15.2	69.1	1568	14	AAE058990	Mutant human creat
38	15.2	69.1	2064	12	AAE014178	Gene encoding asco
39	15.2	69.1	2138	19	AAE64549	M. tuberculosis im
40	15.2	69.1	2138	19	AAE44440	Mycobacterium tube
41	15.2	69.1	2138	20	AAE19350	M. tuberculosis an
42	15.2	69.1	2138	20	AAE19138	M. tuberculosis re
43	15.2	69.1	2343	20	AAE99989	Creatine kinase CK
44	15.2	69.1	2438	21	AAE21626	Human breast and o
45	15.2	69.1	4067	15	AAE63192	Human adeno-associ

ALIGNMENTS

RESULT 1

ID AAT27313 standard; cDNA; 22 BP.

XX AAT27313:

XX 26-NOV-1996 (first entry)

DE T. cruzi epimastigotic PFC100t antigen primer corresp. to bases 1266-87.

KW Antigen: Trypanosoma cruzi; epimastigote; serum; Chagas disease; probe;

KM Primer: PCR; polymerase chain reaction; amplification; antibody; ss.

XX Synthetic.

OS FR2723589-A1.

PN 16-FEB-1996.

XX 12-AUG-1994; 94PR-0010132.

XX 12-AUG-1994; 94PR-0010132.

XX (INMR) BIO MERIEUX.

PA Jolivet M, Lesenechal M, Paranhos-Baccala G;

DR WPI; 1996-190287/20.

PT New nucleic acid encoding Trypanosoma cruzi epimastigotic antigen -

XX useful for diagnosis, monitoring and therapy of Chagas disease

PS Claim 29; Page 37; 55pp; French.

CC The primers AAT27311-5 were used to PCR amplify the sequence encoding a
 CC novel isolated antigenic protein from Trypanosoma cruzi epimastigotes,
 CC designated PTC100t (AAT27310). The primers, derived from the sequences
 CC of a 594 and 1041 bp fragment of PTC100t, amplified the gene as 3
 CC fragments. This primer corresponds to nucleotides 1266-87 of the pTC100t
 CC sequence, derived from the 594 bp fragment. The 594 bp fragment was
 CC isolated from a T. cruzi genomic expression library in lambda gt11, using
 CC a mixture of sera from patients with Chagas disease. It corresponds to
 CC nucleotides 1232-1825 of PTC100t. The 1041 bp fragment was isolated from
 CC a lambda gt10 library using the 594 bp fragment as a probe. The protein,
 CC or antibodies raised against it, can be used in the detection and
 CC monitoring of T. cruzi infection i.e. Chagas disease.

CC Sequence 22 BP; 3 A; 8 C; 7 G; 4 T; 0 other;

Query Match 100.0%; Score 22; DB 17; Length 22;
 Best Local Similarity 100.0%; Pred. No. 0.12; Indels 0; Gaps 0;
 Matches 22; Conservative 0; Mismatches 0;

OY 1 cagcgacgtagctgcgtcct 22
 ||||||||||||||||
 DB 1 cagcgacgtagctgcgtcct 22

RESULT 2

AAx84095
 ID AAx84095 standard; DNA; 22 BP.

AC AAx84095;

DT 27-AUG-1999 (first entry)

DE PCR primer for T. cruzi PTC40 coding sequence.

KW PTC40; Tc40; infection; diagnosis; immune complex; antigenic determinant;
 KM therapy; antibody; PCR primer; ss.

OS Synthetic.

OS Trypanosoma cruzi.

PN MO9929867-A1.

PD 17-JUN-1999.

PF 10-DEC-1998; 98WO-1B01987.

PR 10-DEC-1997; 97US-0988242.

PA (INMR) BIO MERIEUX.

PI Jolivet M, Lesenechal M, Mandrand B, Paranhos-Baccala G;

DR WPI: 1999-394978/33.

PT New Trypanosoma cruzi antigen

PS Disclosure; Page 21; 65pp; English.

CC This sequence is a PCR primer for DNA encoding the Trypanosoma cruzi
 CC PTC40 protein of the invention. The PTC40 antigenic determinant is
 CC useful as a reagent for detection and/or monitoring of Trypanosoma cruzi
 CC infection from samples including blood serum or plasma, urine, saliva, or
 CC tears, by contacting with the sample and detecting an immune complex. The
 CC PTC40 antigenic determinant, the vector, expression cassette, cell or
 CC antibody are useful for treatment or prevention (vaccine) of a
 CC Trypanosoma cruzi infection in a man or animal. Current Trypanosoma cruzi
 CC antigens are obtained from protein fractions of the noninfectious stage
 CC of the parasite, and these do not allow sufficient production of antigens
 CC for use in reliable serological diagnostic tests. The strain to strain
 CC polymorphism reduces reliability of the tests.

Sequence 22 BP; 3 A; 8 C; 7 G; 4 T; 0 other;

Query Match 100.0%; Score 22; DB 20; Length 22;
 Best Local Similarity 100.0%; Pred. No. 0.12; Indels 0; Gaps 0;
 Matches 22; Conservative 0; Mismatches 0;

OY 1 cagcgacgtagctgcgtcct 22
 ||||||||||||||||
 DB 1 cagcgacgtagctgcgtcct 22

RESULT 3

AAAT27310
 ID AAT27310 standard; cDNA; 3402 BP.

AC AAT27310;

DT 26-NOV-1996 (first entry)

DE Trypanosoma cruzi epimastigotic PTC100t antigen gene.

KW Antigen; Trypanosoma cruzi; epimastigote; serum; Chagas disease; probe;
 KM Primer; PCR; polymerase chain reaction; amplification; antibody; ds.

OS Trypanosoma cruzi.

FT Key Location/Qualifiers

FT CDS 266..3013

FT /tag= a
 /product= PTC100t epimastigotic antigen

PN FR2723589-A1.

PD 16-FEB-1996.

PF 12-AUG-1994; 94FR-0010132.

PR 12-AUG-1994; 94FR-0010132.

PA (INMR) BIO MERIEUX.

PI Jolivet M, Lesenechal M, Paranhos-Baccala G;

DR WPI: 1996-190287/20.

DR P-PSDB; AAR91615.

PT New nucleic acid encoding Trypanosoma cruzi epimastigotic antigen -
 useful for diagnosis, monitoring and therapy of Chagas disease

PS Claim 1; Page 24-26; 55pp; French.

CC This is the nucleotide sequence encoding a novel isolated antigenic
 CC protein from Trypanosoma cruzi epimastigotes, designated PTC100t.
 CC The clone Tc50 was isolated from a T. cruzi genomic expression library in
 CC lambda gt11, using a mixture of sera from patients with Chagas disease.
 CC Clone Tc50 contained a 594 bp insert corresp. to nucleotides 1232-1825
 CC of this sequence. The Tc50 sequence was subsequently used to probe a
 CC southern blot of restriction enzyme digested T. cruzi DNA and also screen
 CC a lambda gt10 library to isolate a 1041 bp EcoRI fragment corresp. to
 CC nucleotides 1403-2443 of PTC100t. Primers (AAT27311-5) were synthesised
 CC based on the sequences of the 594 and 1041 bp fragments and used to
 CC amplify the PTC100t clone as 3 fragments from cDNA derived from mRNA
 CC purified from T. cruzi epimastigotes. The protein or antibodies raised
 CC against it can be used in the detection and monitoring of T. cruzi
 CC infection i.e. Chagas disease.

Sequence 3402 BP; 889 A; 818 C; 958 G; 737 T; 0 other;

Query Match 100.0%; Score 22; DB 17; Length 3402;
 Best Local Similarity 100.0%; Pred. No. 0.14; Indels 0; Gaps 0;
 Matches 22; Conservative 0; Mismatches 0;

OY 1 cagccgacggtagctgcctct 22
 |||||||
 DB 1266 cagccgacggtagctgcctct 1287

RESULT 4

AAK84092

ID AAK84092 standard; cDNA; 3402 BP.

XX AAK84092;

XX 27-AUG-1999 (first entry)

DE T. cruzi PTC40 coding sequence.

XX PTC40; Tc40; infection; diagnosis; immune complex; antigenic determinant;

XX therapy; antibody; ds.

XX Trypanosoma cruzi.

XX MO9929867-A1.

XX 17-JUN-1999.

XX 10-DEC-1998; 98MO-1B01987.

XX 10-DEC-1997; 97US-0988242.

XX (INMR) BIO MERIEUX.

XX Jolivet M, Lesenechal M, Mandrand B, Paranhos-Baccala G;

XX MPI: 1999-394978/33.

XX P-PSDB; AAY22124.

XX New Trypanosoma cruzi antigen

XX Claim 1: Page 52-56; 65pp; English.

CC This sequence encodes the Trypanosoma cruzi PTC40 protein of the
 CC invention, and is designated Tc40. The PTC40 antigenic determinant is
 CC useful as a reagent for detection and/or monitoring of Trypanosoma cruzi
 CC infection from samples including blood serum or plasma, urine, saliva, or
 CC tears, by contacting with the sample and detecting an immune complex. The
 CC PTC40 antigenic determinant, the vector, expression cassette, cell or
 CC antibody are useful for treatment or prevention (vaccine) of a
 CC Trypanosoma cruzi infection in a man or animal. Current Trypanosoma cruzi
 CC antigens are obtained from protein fractions of the noninfectious stage
 CC of the parasite, and these do not allow sufficient production of antigens
 CC for use in reliable serological diagnostic tests. The strain to strain
 CC polymorphism reduces reliability of the tests.

SQ Sequence 3402 BP; 888 A; 821 C; 956 G; 737 T; 0 other;

Query Match 100.0%; Score 22; DB 20; Length 3402;

Best Local Similarity 100.0%; Pred. No. 0.14;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cagccgacggtagctgcctct 22
 |||||||
 DB 1266 cagccgacggtagctgcctct 1287

RESULT 5

AAC24763

ID AAC24763 standard; cDNA; 123 BP.

XX AAC24763;

XX 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 28838.

XX Human: 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KM gene therapy; chromosome mapping; ss.
 XX Homo sapiens.
 OS
 XX EP1033401-A2.
 PN
 XX
 XX 06-SEP-2000.
 PD
 XX 21-FEB-2000; 2000EP-0200610.
 PE
 XX 26-FEB-1999; 99US-0122487.
 PR
 XX (CEST) GENSET.

PA (CEST) GENSET.

PI Dumas Malne Edwards J, Duclet A, Giordano J;

XX MPI: 2000-500381/45.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for

XX obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for

XX diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX Claim 1: SEQ ID 28838; 71pp + CD-ROM; English.

CC The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or poly(A) RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors.

SQ Sequence 123 BP; 24 A; 44 C; 35 G; 20 T; 0 other;

Query Match 73.6%; Score 16.2; DB 21; Length 123;

Best Local Similarity 85.7%; Pred. No. 72;

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 cagccgacggtagctgcctct 21
 ||||| || |||||||
 DB 75 cagccgacggtagctgcctct 95

RESULT 6

AAN80616/c

ID AAN80616 standard; DNA; 10475 BP.

XX AAN80616;

XX 21-NOV-1990 (first entry)

DE Genomic clone encoding SAP(Phe).

XX SAP(Val); SAP(Phe); hyaline membrane disease; ss.

XX Key Location/Qualifiers

XX 1054..1120

XX 1431..1538

XX 2060..2131

XX exon

FT		/tag= c	
FT	/label-exon III		
FT	exon	2495..2620	
FT		/tag= d	
FT	/label-exon IV		
FT	exon	3447..3635	
FT		/tag= e	
FT	/label-exon V		
FT	exon	3863..3952	
FT		/tag= f	
FT	/label-exon VI		
FT	exon	5386..5568	
FT		/tag= g	
FT	/label-exon VII		
FT	exon	5742..5887	
FT		/tag= h	
FT	/label-exon VIII		
FT	exon	7122..7202	
FT		/tag= i	
FT	/label-exon IX		
FT	exon	7697..7756	
FT		/tag= j	
FT	/label-exon X		
XX			
PN	W06804324-A.		
PD	16-JUN-1998.		
XX			
PE	03-DEC-1987;	87WO-US03180.	
XX			
PR	03-DEC-1987;	87WO-US03180.	
XX			
PA	03-DEC-1987;	87US-0939206.	
XX			
XX	(ABBO) ABBOTT LABORATORIES.		
PI	Whitsett JA, Fox JL, Pilot-Matias TJ, Meuth JL;		
DR	WPI; 1988-175472/25.		
P-PSDB:	AAP82934.		
XX			
PT	pulmonary hydrophobic surfactant-associated proteins - useful for		
PT	normalising pulmonary surface tension.		
XX			
PS	Disclosure: ; pp; English.		
XX			
CC	This genomic clone encodes the pulmonary hydrophobic surfactant-		
CC	associated protein SAP(Phe). Following processing to remove the		
CC	introns the prod. is obt'd. in a high yield. When SAP(Phe) is		
CC	combined with a lipid it can be used to reduce or maintain		
CC	normal pulmonary surface tension in the alveoli of animals		
CC	(esp. humans) and therefore can be used in the treatment of		
CC	hyaline membrane disease in premature infants.		
CC	May also be used to deliver substances to respiratory epithelial cells.		
CC	See also AAP80572-96, AAN80613, AAN80615, AAN80617 and AAN82412-13.		
XX			
SO	Sequence 10475 BP; 2416 A; 3063 C; 2940 G; 2056 T; 0 other;		
OY			
Db	Query Match	73.6%;	Score 16.2; DB 9; Length 10475;
	Best Local Similarity	85.7%;	Pred. No. 83;
	Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0		
	2 accgcagctagctcgtccct 22		
	AgCCCAAGCTTGTGTCGTGTCCT 8349		
RESULT 7			
AAAN80643/c			
ID AAAN80643 standard; DNA; 10475 BP.			
XX			
AC AAAN80643;			
XX			

DT	03-FEB-1991	(first entry)	
XX			
DE	Sequence of human pulmonary hydrophobic surfactant-associated protein		
DE	(SAP) (Phe) genomic clone.		
XX			
KW	Human pulmonary hydrophobic surfactant-associated protein (SAP) (Phe)		
RW	hyaline membrane disease (HMD); therapy; ss.		
XX			
OS	Homo sapiens.		
XX			
FT	Key	Location/Qualifiers	
FT	exon	1054..1119	
FT		/*tag= a	
FT		/note="AAP80193"	
FT	exon	1433..1558	
FT		/codon_start=1433	
FT		/*tag= b	
FT	exon	/note="AAP80194"	
FT		2059..2130	
FT		/*tag= c	
FT	exon	/note="AAN80195"	
FT		2494..2619	
FT		/*tag= d	
FT	exon	/note="AAP81070"	
FT		3445..3633	
FT		/*tag= e	
FT	exon	/note="AAP81074"	
FT		3861..3950	
FT		/*tag= f	
FT	exon	/note="AAP81075"	
FT		5384..5567	
FT		/*tag= g	
FT	exon	/note="AAP81076"	
FT		5742..5885	
FT		/*tag= h	
FT	exon	/codon_start=5742	
FT		/note="AAP81077"	
FT		7120..7200	
FT		/*tag= i	
FT	exon	/note="AAP81078"	
FT		7695..7754	
FT		/*tag= j	
FT		/note="AAP81079"	
XX			
PN	W08803170-A.		
PD			
PD	05-MAY-1988.		
XX			
PE	02-OCT-1987;	87WO-US02536.	
XX			
XX	01-OCT-1987;	87US-0101680.	
PR	08-DEC-1986;	86US-0939206.	
PR	10-JUN-1987;	87US-0060719.	
XX			
PA	(WHIT/) WHITSETT JA.		
XX			
PI	Whitsett JA, Fox JL, Pilot-Matias TJ, Meuth JL, Sarin VK;		
XX			
DR	WPI: 1988-133244/19.		
DR	P-PSDB; AAP80193, AAP80194, AAP80195, 081070, AAP81074, AAP81075,		
DR	AAP81076, AAP81077, AAP81078, AAP81079.		
XX			
PT	Pulmonary hydrophobic surfactant-associated proteins -		
PT	used with lipid(s) to treat and prevent hyaline membrane disease		
XX	and similar syndromes		
PS	Example; Fig 8B-1 - 8B-24; 144pp; English.		
XX			
CC	The sequence is derived from human embryonic kidney cell genomic		
CC	library. SAP (Val) and SAP (Phe), when combined with lipids, have		
CC	significant pulmonary biophysical surfactant activity that may be		
CC	utilised to treat and prevent hyaline membrane disease (HMD) and		
CC	other syndromes associated with lack or insufficient amts. of natural		

DT 27-JAN-1999 (first entry)
 XX
 DE M. tuberculosis immunogenic polypeptide Tbh-8-2 DNA.
 XX
 XX Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
 KM vaccine; pharmaceutical; infection; diagnosis; ss.
 OS
 OS Mycobacterium tuberculosis.
 XX
 PN WO9816646-A2.
 XX
 PD 23-APR-1998.
 XX
 PF 07-OCT-1997; 97WO-US18293.
 XX
 PR 13-MAR-1997; 97US-0818112.
 PR 11-OCT-1996; 96US-0730510.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
 PI Reed SG, Skelky YAW, Twardzik DR, Vedvick TS;
 XX
 DR WPI; 1998-261042/23.
 XX
 PS Example 3b; Page 126; 230pp; English.
 CC This sequence encodes an immunogenic portion of a soluble Mycobacterium
 CC tuberculosis (MT) antigen which can be used in a method for inducing
 CC protective immunity against tuberculosis (TB). This sequence can be
 CC formulated into vaccines and/or pharmaceutical compositions for
 CC immunising against M. tuberculosis infection or may be used for the
 CC diagnosis of tuberculosis.
 CC
 CC Sequence 282 BP; 56 A; 79 C; 86 G; 54 T; 7 other;
 XX
 SQ

Query Match 71.8%; Score 15.8; DB 19; Length 282;
 Best Local Similarity 85.0%; Pred. No. 1.1e+02;
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 cagccgacggtagctgcgc 20
 |||||
 Db 136 cagccgacggtgctgcgc 155

RESULT 11
 AAV44394
 ID AAV44394 standard; DNA; 282 BP.
 XX
 AC AAV44394;
 XX
 DT 09-NOV-1998 (first entry)
 XX
 DE Mycobacterium tuberculosis antigen Tbh-8-2 DNA.
 XX
 KM Tuberculosis; infection; diagnosis; antigen; Tbh-8-2; ss.
 OS
 OS Mycobacterium tuberculosis strain H37Rv.
 XX
 PN WO9816645-A2.
 XX
 PD 23-APR-1998.
 XX
 PF 07-OCT-1997; 97WO-US18214.
 XX
 PR 13-MAR-1997; 97US-0818111.
 PR 11-OCT-1996; 96US-0729622.
 XX

PA (CORI-) CORIXA CORP.
 XX
 PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
 PI Reed SG, Skelky YAW, Twardzik DR, Vedvick TS;
 XX
 DR WPI; 1998-251292/22.
 XX
 PS New isolated Mycobacterium tuberculosis polypeptides and DNA - used
 PT to develop products for the detection of M. tuberculosis infection
 PT and diagnosis of tuberculosis
 XX
 PS Example 3; Page 131; 250pp; English.
 CC This DNA sequence codes for an antigenic portion of Mycobacterium
 CC tuberculosis antigen Tbh-8-2. It was isolated from a M.
 CC tuberculosis strain H37Rv expression library using sera from
 CC patients having pulmonary or pleural tuberculosis, and is a partial
 CC clone of Tbh-8 (see AAV44370). The invention relates to compositions
 CC and methods for diagnosing tuberculosis. It provides polypeptides
 CC (see AAV64291-W64379) comprising an antigenic portion of a soluble M.
 CC tuberculosis antigen, or an immunogenic portion of an M.
 CC tuberculosis antigen, as well as DNA sequences encoding such
 CC polypeptides, recombinant expression vectors and transformed or
 CC transfected host cells. Also claimed are methods and diagnostic
 CC kits for detecting M. tuberculosis infection in a patient using
 CC these polypeptides, antibodies or oligonucleotide probes and
 CC primers, for the diagnosis of tuberculosis.
 XX
 SQ Sequence 282 BP; 56 A; 79 C; 86 G; 54 T; 7 other;

Query Match 71.8%; Score 15.8; DB 19; Length 282;
 Best Local Similarity 85.0%; Pred. No. 1.1e+02;
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 cagccgacggtagctgcgc 20
 |||||
 Db 136 cagccgacggtgctgcgc 155

RESULT 12
 AA219304
 ID AA219304 standard; DNA; 282 BP.
 XX
 AC AA219304;
 XX
 DT 05-NOV-1999 (first entry)
 XX
 DE M. tuberculosis antigen Tbh-8-2 DNA sequence.
 XX
 KM Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
 KM immunotherapy; diagnosis; immunisation; vaccine; infection;
 KM immune response; skin test; ss.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN WO9942076-A2.
 XX
 PD 26-AUG-1999.
 XX
 PF 17-FEB-1999; 99WO-US03268.
 XX
 PR 05-MAY-1998; 98US-0072967.
 PR 18-FEB-1998; 98US-0025197.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
 PI Lodes MJ, Reed SG, Skelky YAW, Twardzik DR, Vedvick TS;
 XX
 DR WPI; 1999-527409/44.
 XX
 PT New antigens from Mycobacterium tuberculosis useful in diagnostic

PT skin tests and protective or therapeutic vaccines or compositions
 XX
 PS Example 3; Page 120; 299pp; English.
 XX
 CC The present invention describes polypeptides comprising an immunogenic
 CC part of a Mycobacterium tuberculosis antigen (Ag). Also described
 CC are vaccines and fusion protein containing M. tuberculosis Ag's.
 CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and
 CC other polypeptides fragments, can be used in pharmaceutical compositions
 CC or vaccines to generate a protective or therapeutic immune response to
 CC M. tuberculosis and as reagents in skin tests for diagnosis of
 CC tuberculosis. Ag can induce proliferation of, or cytokine secretion
 CC by, T, B or natural killer cells and/or macrophages in
 CC tuberculosis-immune subjects. AA19249 to AA219460 and AA19083 to
 CC AA19225 are used in the exemplification of the present invention.
 XX
 SQ Sequence 282 BP; 56 A; 79 C; 86 G; 54 T; 7 other;

Query Match 71.8%; Score 15.8; DB 20; Length 282;
 Best Local Similarity 85.0%; Pred. No. 1.1e+02;
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 cagcgacgagtgctgcgc 20
 |||||
 DB 136 cagcgacgagtgctgcgc 155

RESULT 13
 AA219092
 ID AA219092 standard; DNA; 282 BP.
 XX
 AC AA219092;
 XX
 DT 05-NOV-1999 (first entry)
 XX
 DE M. tuberculosis recombinant antigen DNA encoding Tbh-8-2.
 XX
 KW Antigen; diagnosis; detection; infection; antibody; immunisation;
 KW vaccine; immunity; ss.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PA WO9942118-A2.
 XX
 PD 26-AUG-1999.
 XX
 PF 17-FEB-1999; 99MO-US03265.
 XX
 PR 05-MAY-1998; 98US-0072596.
 PR 18-FEB-1998; 98US-0024753.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
 PI Lodes MJ, Reed SG, Skelky YAM, Twardzik DR, Vedvick TS;
 XX
 DR WPI; 1999-527416/44.
 XX
 PT New polypeptide comprising antigenic portions of M. tuberculosis
 PT
 PS Example 3; Page 165-166; 323pp; English.
 XX
 CC This invention describes novel recombinant antigens and their encoding
 CC nucleic acids derived from Mycobacterium tuberculosis. The novel
 CC polypeptides are useful for detecting M. tuberculosis infection in a
 CC biological sample by detecting antibodies which bind with the
 CC polypeptides, and are useful as vaccines for immunizing against
 CC M. tuberculosis infection. The new detection methods are needed as
 CC current vaccination strategies do not provide 100% immunity.
 CC
 SQ Sequence 282 BP; 56 A; 79 C; 86 G; 54 T; 7 other;

Query Match 71.8%; Score 15.8; DB 20; Length 282;
 Best Local Similarity 85.0%; Pred. No. 1.1e+02;
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 cagcgacgagtgctgcgc 20
 |||||
 DB 136 cagcgacgagtgctgcgc 155

RESULT 14
 AAF68508/c
 ID AAF68508 standard; cDNA; 404 BP.
 XX
 AC AAF68508;
 XX
 DT 12-APR-2001 (first entry)
 XX
 DE Human lung tumour protein related nucleotide sequence SPQ ID NO:443.
 XX
 KW Human; lung cancer; lung tumour; lung tumour protein; gene therapy;
 KW lung cancer antigen; lung tumour-specific antigen; diagnosis; vaccine;
 KW cytosolic; antisense inhibition; ss.
 XX
 OS Homo sapiens.
 XX
 PA WO200100828-A2.
 XX
 PD 04-JAN-2001.
 XX
 PF 30-JUN-2000; 2000WO-US18061.
 XX
 PR 30-JUN-1999; 99US-0346492.
 PR 15-OCT-1999; 99US-0419356.
 PR 17-DEC-1999; 99US-0466867.
 PR 30-DEC-1999; 99US-0476300.
 PR 06-MAR-2000; 2000US-0519642.
 PR 22-MAR-2000; 2000US-0533077.
 PR 10-APR-2000; 2000US-0546259.
 PR 27-APR-2000; 2000US-0560406.
 PR 05-JUN-2000; 2000US-0589184.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Wang T, Bangur CS, Lodes MJ, Fanger GR, Vedvick TS, Carter D;
 PI Retter MW, Mannion J;
 XX
 DR WPI; 2001-071488/08.
 XX
 PT Lung tumor-associated proteins and the nucleic acids that encode them,
 PT useful for preventing, diagnosing and treating lung cancer -
 XX
 PS Claim 4; Page 302; 436pp; English.
 XX

The present invention describes immunogenic portions of lung tumour-associated proteins (I) and the nucleic acids (NAs) that encode them. (I) have cytosolic activity and can be used in gene therapy, antisense inhibition and in vaccines. The NAs and the lung tumour-associated proteins they encode may be used in the prevention, treatment and diagnosis of diseases associated with their inappropriate expression, especially lung cancers. For example, the NAs may be administered to treat diseases by rectifying mutations or deletions in a patient's genome that affect the activity of the protein by expressing inactive proteins or to supplement the patient's own production of (I). Additionally, the NAs may be used to produce the lung-tumour associated protein, according to standard recombinant DNA methodology. Conversely, antisense NA molecules may be administered to down regulate protein expression by binding with the cells own genes and preventing their expression. The NA and complementary sequences may also be used as DNA probes in diagnostic assays to detect and quantitate the presence of similar NA sequences in samples, and hence which patients may be in need of treatment for lung cancer. The (I) may be used as antigens in the production of antibodies and in assays to identify modulators (agonists and antagonists) of the

CC expression and activity of the protein. AAF68083 to AAF68878 and
CC AAB76848 to AAB76878 represent human lung tumour protein related
CC nucleotide and protein sequences which are used in the exemplification
CC of the present invention.

XX
SQ Sequence 404 BP; 87 A; 112 C; 112 G; 85 T; 8 other;

Query Match 71.8%; Score 15.8; DB 22; Length 404;
Best Local Similarity 81.0%; Pred. No. 1.2e+02;

Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 agccgacgtagctgcgtcct 22
||||| ||||| ||||| |||||

Db 316 AGCCGCGTGGTGTCTCTCT 296

RESULT 15

AAA37040
ID AAA37040 standard; cDNA; 1661 BP.

XX AAA37040;

XX 08-AUG-2000 (first entry)

DE Human PRO1772 (UNC834) cDNA sequence SEQ ID NO:62.

XX Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;

KW transmembrane; secretion; immunoadhesion; pharmaceutical; screening;

KW SS.

XX Homo sapiens.

PN WO200012708-A2.

XX 09-MAR-2000.

PD 09-MAR-2000.

XX 09-MAR-2000.

XX 09-MAR-2000.

XX 09-MAR-2000.

XX 09-MAR-2000.

XX 09-MAR-2000.

XX 09-MAR-2000.

XX 09-MAR-2000.

XX 09-MAR-2000.

XX 09-MAR-2000.

XX 09-MAR-2000.

XX 09-MAR-2000.

XX 09-MAR-2000.

XX 09-MAR-2000.

XX 09-MAR-2000.

XX 09-MAR-2000.

XX 09-MAR-2000.

XX 09-MAR-2000.

XX 09-MAR-2000.

XX 09-MAR-2000.

PR 18-SEP-1998; 98US-0100849.

PR 18-SEP-1998; 98US-0101014.

PR 18-SEP-1998; 98US-0101068.

PR 18-SEP-1998; 98US-0101071.

PR 22-SEP-1998; 98US-0101279.

PR 23-SEP-1998; 98US-0101471.

PR 23-SEP-1998; 98US-0101472.

PR 23-SEP-1998; 98US-0101474.

PR 23-SEP-1998; 98US-0101475.

PR 23-SEP-1998; 98US-0101476.

PR 23-SEP-1998; 98US-0101477.

PR 23-SEP-1998; 98US-0101479.

PR 24-SEP-1998; 98US-0101738.

PR 24-SEP-1998; 98US-0101741.

PR 24-SEP-1998; 98US-0101743.

PR 24-SEP-1998; 98US-0101915.

PR 24-SEP-1998; 98US-0101916.

PR 29-SEP-1998; 98US-0102207.

PR 29-SEP-1998; 98US-0102240.

PR 29-SEP-1998; 98US-0102307.

PR 29-SEP-1998; 98US-0102330.

PR 29-SEP-1998; 98US-0102331.

PR 30-SEP-1998; 98US-0102484.

PR 30-SEP-1998; 98US-0102487.

PR 30-SEP-1998; 98US-0102570.

PR 30-SEP-1998; 98US-0102571.

PR 01-OCT-1998; 98US-0102684.

PR 01-OCT-1998; 98US-0102687.

PR 02-OCT-1998; 98US-0102685.

PR 06-OCT-1998; 98US-0103258.

PR 06-OCT-1998; 98US-0103449.

PR 07-OCT-1998; 98US-0103314.

PR 07-OCT-1998; 98US-0103315.

PR 07-OCT-1998; 98US-0103328.

PR 07-OCT-1998; 98US-0103395.

PR 07-OCT-1998; 98US-0103396.

PR 07-OCT-1998; 98US-0103401.

PR 08-OCT-1998; 98US-0103633.

PR 08-OCT-1998; 98US-0103678.

PR 08-OCT-1998; 98US-0103679.

PR 14-OCT-1998; 98US-0103711.

PR 20-OCT-1998; 98US-0104257.

PR 20-OCT-1998; 98US-0104987.

PR 20-OCT-1998; 98US-0105000.

PR 20-OCT-1998; 98US-0105002.

PR 21-OCT-1998; 98US-0105104.

PR 22-OCT-1998; 98US-0105169.

PR 22-OCT-1998; 98US-0105266.

PR 26-OCT-1998; 98US-0105693.

PR 26-OCT-1998; 98US-0105694.

PR 27-OCT-1998; 98US-0105807.

PR 27-OCT-1998; 98US-0105882.

PR 27-OCT-1998; 98US-0105881.

PR 27-OCT-1998; 98US-0106023.

PR 28-OCT-1998; 98US-0106029.

PR 28-OCT-1998; 98US-0106032.

PR 28-OCT-1998; 98US-0106037.

PR 28-OCT-1998; 98US-0106033.

PR 28-OCT-1998; 98US-0106178.

PR 29-OCT-1998; 98US-0106248.

PR 29-OCT-1998; 98US-0106384.

PR 29-OCT-1998; 98US-0106500.

PR 30-OCT-1998; 98US-0106464.

PR 30-OCT-1998; 98US-0106856.

PR 03-NOV-1998; 98US-0106902.

PR 03-NOV-1998; 98US-0106905.

PR 03-NOV-1998; 98US-0106919.

PR 03-NOV-1998; 98US-0106932.

PR 03-NOV-1998; 98US-0106934.

PR 10-NOV-1998; 98US-0107783.

PR 17-NOV-1998; 98US-0108775.

PR 17-NOV-1998; 98US-0108779.

PR 17-NOV-1998; 98US-0108787.
 PR 17-NOV-1998; 98US-0108788.
 PR 17-NOV-1998; 98US-0108801.
 PR 17-NOV-1998; 98US-0108802.
 PR 17-NOV-1998; 98US-0108806.
 PR 17-NOV-1998; 98US-0108807.
 PR 17-NOV-1998; 98US-0108867.
 PR 17-NOV-1998; 98US-0108925.
 PR 18-NOV-1998; 98US-0108848.
 PR 18-NOV-1998; 98US-0108849.
 PR 18-NOV-1998; 98US-0108850.
 PR 18-NOV-1998; 98US-0108851.
 PR 18-NOV-1998; 98US-0108852.
 PR 18-NOV-1998; 98US-0108858.
 PR 18-NOV-1998; 98US-0108904.

XX (GETH) GENENTECH INC.

PA Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;
 XX

PI WPI: 2000-237871/20.
 XX

DR P-PSDB; AAY99358.
 DR

XX New mammalian DNA sequences encoding transmembrane, receptor or
 PT secreted PRO polypeptides, useful for screening of potential peptide or
 PT small molecule inhibitors of the relevant receptor/ligand interactions
 XX

PS Claim 2; Fig 37; 773pp; English.

XX AAA37022 to AAA37144 encode the new isolated human transmembrane,
 CC receptor or secreted PRO polypeptides given in AAY9340 to AAY9462. The
 CC transmembrane and receptor PRO proteins can be used for screening of
 CC potential peptide or small molecule inhibitors of the relevant
 CC receptor/ligand interactions. The polypeptides and nucleotide sequences
 CC encoding them have various industrial applications, including uses as
 CC pharmaceutical and diagnostic agents. AAA37145 to AAA37330 represent
 CC PCR primers and hybridisation probes used in the isolation of the PRO
 CC polypeptides from the present invention.
 XX

SO Sequence 1661 BP; 334 A; 519 C; 457 G; 349 T; 2 other;

Query Match 70.9%; Score 15.6; DB 21; Length 1661;

Best Local Similarity 81.8%; Pred. No. 1.5e+02;

Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 OY 1 cagccgacgtagctgcgtcct 22
 ||||| ||||| ||||| |
 Db 128 cagccgacgtagctgcgtcgt 149

Search completed: September 21, 2001, 02:42:41
 Job time: 11457 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compgen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2001, 23:28:03 ; Search time 4309.39 Seconds
(without alignments)
48.258 Million cell updates/sec

Title: US-09-138-735-9

Perfect score: 22
Sequence: 1 cagccgacgtagctgcgtcct 22

Scoring table: IDENTITY-NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_esthum29:*
64: em_esthum30:*
65: em_esthum31:*
66: em_esthum32:*
67: em_esthum33:*
68: em_esthum34:*
69: em_esthum35:*
70: em_esthum36:*
71: em_esthum37:*
72: em_esthum38:*
73: em_esthum39:*
74: em_esthum40:*
75: em_esthum41:*
76: em_esthum42:*
77: em_esthum43:*
78: em_esthum44:*
79: em_esthum45:*
80: em_esthum46:*
81: em_esthum47:*
82: em_esthum48:*
83: em_esthum49:*
84: em_esthum50:*
85: em_esthum51:*
86: em_esthum52:*
87: em_esthum53:*
88: em_esthum54:*
89: em_esthum55:*
90: em_esthum56:*
91: em_esthum57:*
92: em_esthum58:*
93: em_esthum59:*
94: em_esthum60:*
95: em_esthum61:*
96: em_esthum62:*
97: em_esthum63:*
98: em_esthum64:*
99: em_esthum65:*
100: em_esthum66:*
101: em_esthum67:*
102: em_esthum68:*
103: em_esthum69:*
104: em_esthum70:*
105: em_esthum71:*
106: em_esthum72:*
107: em_esthum73:*
108: em_esthum74:*
109: em_esthum75:*
110: em_esthum76:*
111: em_esthum77:*
112: em_esthum78:*
113: em_esthum79:*
114: em_esthum80:*
115: em_esthum81:*
116: em_esthum82:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	17.4	79.1	461	175	BG270722	BG270722 la97h10.y
2	17.2	78.2	177	21	AI96394	AI96394 sb04b07.y
3	17.2	78.2	205	114	AW306389	AW306389 se50a06.y
4	17.2	78.2	452	234	AO872227	AO872227 nbe00470
5	17.2	78.2	481	152	BG365834	BG365834 HVSME1000
6	17.2	78.2	540	19	AI392116	AI392116 NCC2B4T3
7	17.2	78.2	594	121	AM681895	AM681895 CM3-CT034
8	17.2	78.2	744	14	AA950463	AA950463 LD30267.5
9	17.2	78.2	910	153	BG414942	BG414942 HVSMEK000
C 10	17.2	78.2	958	142	BE962535	BE962535 601655848
C 11	17.2	78.2	993	222	CNS050R7	AL447032 Tetradon
C 12	16.8	76.4	259	113	AM265763	AM265763 xp81e09.x
C 13	16.8	76.4	281	170	BF825637	BF825637 RC0-HN002
C 14	16.8	76.4	332	107	AU032424	AU032424 AU032424
C 15	16.4	74.5	928	172	BF978349	BF978349 602148933
C 16	16.2	73.6	209	122	AM922321	AM922321 DG1.17.E0
C 17	16.2	73.6	246	116	AM445732	AM445732 82893 MAR
C 18	16.2	73.6	295	116	AM478058	AM478058 18366 MAR
C 19	16.2	73.6	338	222	FR0031088	AL027457 Fugu tubr
C 20	16.2	73.6	357	224	AO073528	AO073528 EP(2)2366
C 21	16.2	73.6	366	116	AM435588	AM435588 74113 MAR
C 22	16.2	73.6	395	158	HA48232	HA48232 YG69A01.s1
C 23	16.2	73.6	402	116	AM463080	AM463080 BP230011A
C 24	16.2	73.6	403	116	AM447281	AM447281 88385 MAR
C 25	16.2	73.6	428	143	BF044716	BF044716 BP250001B
C 26	16.2	73.6	451	139	BE758397	BE758397 SWYACAL10
C 27	16.2	73.6	455	162	BE050606	BE050606 za67g01.b
C 28	16.2	73.6	457	16	AI105748	AI105748 SWACAC25
C 29	16.2	73.6	458	136	BE499187	BE499187 WHE0970-C
C 30	16.2	73.6	460	162	BE050607	BE050607 za67g01.g
C 31	16.2	73.6	464	117	AW510932	AW510932 hda1d12.x
C 32	16.2	73.6	471	22	AI566236	AI566236 tq7db06.x
C 33	16.2	73.6	472	143	BF043469	BF043469 BP250025B
C 34	16.2	73.6	477	147	BF358274	BF358274 PM1-EN000
C 35	16.2	73.6	485	188	R61321	R61321 YH07e07.s1
C 36	16.2	73.6	490	114	AM309353	AM309353 s116c09.x
C 37	16.2	73.6	497	111	BF909469	BF909469 IL2-UT007
C 38	16.2	73.6	520	143	BF042379	BF042379 BP250022A
C 39	16.2	73.6	521	4	AA240489	AA240489 mw32b07.r
C 40	16.2	73.6	521	220	CNS02LDP	AL202678 Tetradon
C 41	16.2	73.6	526	139	BE757430	BE757430 211865 MA
C 42	16.2	73.6	533	104	AI943345	AI943345 fc79a04.y
C 43	16.2	73.6	538	19	AI403626	AI403626 GH23107.5
C 44	16.2	73.6	540	119	AM657127	AM657127 109776 MA
C 45	16.2	73.6	540	119	AM657127	AM657127 109776 MA

ALIGNMENTS

RESULT 1
LOCUS BG270722/c
DEFINITION la97h10.y1 Melton Mouse Adult Pancreas 1 Mus musculus cDNA 5' similar to SW:TR3_PAT P08426 TRYPSINOGEN III, CATIONIC PRECURSOR
; mRNA sequence.
SESSION BG270722
BG270722.1 GI:12978157
EST 20-FEB-2001
MUS musculus

REFERENCE 1. (bases 1 to 461)
Melton, D., Meadows, A., Clifton, S., Hillier, L., Marra, M., Page, D., Wylie, T., Martin, J., Blistein, A., Schmitt, A., Theising, B., Ritter, E., Nonko, I., Bennett, J., Cardenas, M., Gibbons, M., McCann, R., Cole

TITLE
JOURNAL
COMMENT

R., Tsagaris, H., Williams, T., Jackson, Y., and Bowers, Y.
Washu-Harvard Pancreas EST Project
Unpublished (2000)
Contact: Douglas Melton
Washu-Harvard Pancreas EST Project
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biochem.harvard.edu, meadows@fas.harvard.edu
Library was constructed by Dr. Douglas Melton DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: Juliana Brown
(brown@fas.harvard.edu)
Seq primer: -40RP from Gibco
High quality sequence stop: 445.
Location/Qualifiers
1. 461
/organism="Mus musculus"
/strain="ICR"
/db_xref="taxon:10090"
/clone_lib="Melton Mouse Adult Pancreas 1"
/sex="Male"
/tissue_type="Total Pancreas"
/dev_stage="Adult"
/lab_host="TOP10"
/note="Organ: Pancreas; Vector: pZero-2; Site: 1; Not I; Site: 2; Xho I; Library constructed using Superscript Plasmid Library kit (Life Technologies). cDNA made by oligo-dT priming. Xho I site destroyed during cloning. Size-selected by column fractionation. Primary library, unamplified."
BASE COUNT 92 a 118 c 158 g 84 t 9 others
ORIGIN

Query Match 79.1%; Score 17.4; DB 175; Length 461;
Best Local Similarity 85.7%; Pred. No. 3e+02; Mismatches 18; Conservative 0; Indels 0; Gaps 0;

1 cagccgacggtacgtcgc 21
|||||
Db 202 CAGCCGACGTGACGTGCTCC 182

RESULT 2
LOCUS AI496394
DEFINITION sb04b07.y1 Gm-c1004 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
Gm-c1004-7790 5', mRNA sequence.
SESSION AI496394
AI496394.1 GI:4397397
EST
KEYWORDS
SOURCE soybean.
ORGANISM Glycine max

REFERENCE 1 (bases 1 to 177)
Shoemaker, R., Kelm, P., Vodka, L., Erpelting, J., Correll, V., Khanna
A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,
Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers
Y., Person, B., Waller, T., Gibbons, M., Page, D., Harvey, N., Schurk
R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann
R., Waterston, R. and Wilson, R.
Public Soybean EST Project
Unpublished (1999)

CONTACT: Shoemaker R./Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800

Fax: 314 286 1810
Email: estevenson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134 For further information
call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or
info@genomesystems.com web site: www.genomesystems.com
Seq primer: -40RP from Gibco.

FEATURES

source

Location/Qualifiers
1. 177
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1004-7790"
/clone_1b="Gm-c1004"
/tissue_type="root"
/lab_host="XL10-Gold"
/note="Vector: pBluescript II Xr; Site:1: EcoRI; Site:2:
XhoI; Root cDNA. The mRNA was isolated from entire roots
of 8 day old 'Williams' seedlings which were propagated on
paper towels with distilled water. StrataGene's cDNA
Synthesis Kit (catalog #200401) was used to synthesize the
cDNA. First-strand synthesis was performed with 5-methyl
dCTP, hence the ligated cDNA is hemimethylated.
StrataGene's first-strand synthesis primer was used
[GAGAGAGAGAGAGAGAGACTACTCTGAC(T)18]. After
second-strand synthesis, the cDNA ends were 'polished',
with Clontech Pfu DNA polymerase, ligated to EcoRI adapters,
and phosphorylated. The XhoI site within the first-strand
synthesis primer was restricted by digestion with XhoI;
all XhoI sites in the cDNA would be protected by their
hemimethylated status. The cDNA constructs were
size-fractionated with a 500bp cutoff, using GibcoBRL Life
Technologies' cDNA size fractionation column. The column
eluent was then ligated into StrataGene's pBluescript II
Xr Predigested vector (pBluescript II SK(+)) that had been
digested with EcoRI and XhoI, and phosphorylated. Both
the white and blue colonies appear to contain recombinant
plasmids with cDNA inserts. Blue colonies (96-15) have been
sequenced, and possess putative cDNA inserts. This library
was constructed by Dr. Paul Keim, a Virginia H. Coryell,
Department of Biology, Box5640, Northern Arizona
University, Flagstaff, AZ 86011, Phone: 520-523-1078 (Dr.
Paul Keim), 520-523-1372 (Virginia H. Coryell), Fax:
520-523-7500, email: paul.keim@na.u.edu,
virginia.coryell@na.u.edu"

BASE COUNT 28 a 62 c 25 g 62 t
ORIGIN

Query Match 78.2%; Score 17.2; DB 21; Length 177;
Best Local Similarity 86.4%; Pred. No. 3.8e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 cagccgacgtagctgcgtct 22
Db 150 CAGCCGACGTCGTCCTCT 171

RESULT 3
LOCUS 205 bp mRNA EST 20-JAN-2000
DEFINITION 9650a06.y1 Gm-c1017 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
Gm-c1017-2315 5', mRNA sequence.
ACCESSION AM306389
VERSION AM306389.1 GI:6718742
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eutrosids; Fabales; Fabaceae; Papilionoideae; Glycine.
REFERENCE 1 (bases 1 to 205)
AUTHORS Shoemaker, R., Keim, P., Vodka, L., Erpelting, J., Coryell, V., Khanna

A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,
Wyllie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers
J., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk
R., Rutter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann
R., Waterston, R., and Wilson, R.
Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: estevenson.wustl.edu

TITLE

JOURNAL

COMMENT

FEATURES

source

Location/Qualifiers
1. 205
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1017-2315"
/clone_1b="Gm-c1017"
/tissue_type="vegetable buds of field grown plants"
/lab_host="XL10-Gold"
/note="Vector: pBluescript II Xr; Site:1: EcoRI; Site:2:
XhoI; This cDNA library was constructed from mRNA isolated
from vegetable buds of field grown plants. The cDNA
library was prepared using the StrataGene pBluescript II
Xr library construction kit. Complementary DNA was
synthesized from mRNA using a primer consisting of a poly
(dT) sequence with a XhoI restriction site. EcoRI adapters
were ligated to the blunt-ended cDNA fragments followed by
XhoI digestion. The cDNA fragments were directionally
cloned into the EcoRI-XhoI restriction site of the
pBluescript vector. The ligated cDNA fragments were
transformed into XL10-Gold host cells. This library was
constructed by Dr. Randy Shoemaker and Dr. John
Erpelting."

BASE COUNT 59 a 47 c 41 g 58 t
ORIGIN

Query Match 78.2%; Score 17.2; DB 114; Length 205;
Best Local Similarity 86.4%; Pred. No. 3.8e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 cagccgacgtagctgcgtct 22
Db 31 CAGCCGACGTCGTCCTCT 52

RESULT 4
LOCUS 452 bp DNA GSS 03-NOV-1999
DEFINITION nbe0047019r CUGI Rice BAC Library (EcoRI) Oryza sativa genomic
clone nbe0047019r, DNA sequence.
ACCESSION AO872227
VERSION AO872227.1 GI:6222678
KEYWORDS GSS.
SOURCE Oryza sativa.
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE 1 (bases 1 to 452)
AUTHORS Wing, R.A. and Dean, R.A.
TITLE A BAC End Sequencing Framework to Sequence the Rice Genome
JOURNAL Unpublished (1998)
COMMENT Contact: Wing RA

/clone="NC2B4"
/sex="Mating type A"
/issue_type="Conidia"
/dev_stage="germinating conidia"
/note="mRNA isolated from germinating conidia, grown in 1x Vogel's, 2% sucrose for 4.5 hours. cDNA directionally cloned into pBluescript SK(-) using the uni-ZAP XR vector system (Stratagene, La Jolla, CA)."

BASE COUNT 117 a 172 c 130 g 121 t

ORIGIN

Query Match 78.2%; Score 17.2; DB 19; Length 540;
Best Local Similarity 86.4%; Pred. No. 3.7e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 cagccgacgtagctgcgtct 22
|||||
Db 309 CTGCCGACGGCTGCTGCTCT 330

RESULT 7
AM861895 594 bp mRNA EST 19-MAY-2000
LOCUS AM861895
DEFINITION CM3-CT0341-170200-093-f02 CT0341 Homo sapiens cDNA, mRNA sequence.
ACCESSION AM861895
VERSION AM861895.1 GI:7957588
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 594)
Dias Neto, E., Garcia Correia, R., Verjovski-Almeida, S., Brites, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l1=62-CH3-CT0341-170200-093-f02&ts=2000-02-17&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 592.
Location/Qualifiers
1. 594

FEATURES
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CT0341"
/dev_stage="Adult"

/note="Organ: colon; Vector: puc18; Site1: Sma1; Site2: Sma1; A mini-library was made by cloning products derived from ORFESTS PCR (U.S. Letters Patent Application No. 196 /16 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
SE COUNT 134 a 161 c 162 g 137 t
GIN

Query Match 78.2%; Score 17.2; DB 121; Length 594;
Best Local Similarity 86.4%; Pred. No. 3.7e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 cagccgacgtagctgcgtct 22
|||||
Db 157 CAGCCGACGGCTGCTGCTCT 178

RESULT 8
AA950463 744 bp mRNA EST 25-NOV-1998
LOCUS AA950463
DEFINITION LD30267.5prime LD Drosophila melanogaster embryo pot2 Drosophila melanogaster cDNA clone LD30267 5prime, mRNA sequence.
ACCESSION AA950463
VERSION AA950463.1 GI:3112700
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster

REFERENCE
AUTHORS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephygroidae; Drosophilidae; Drosophila.
1 (bases 1 to 744)
Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C., Brokstein, P., Lewis, S. and Rubin, G.M.
BDGP/HMI Drosophila EST Project
Unpublished (1997)
Contact: Harvey, D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 LSA, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: 302 row: F column: 7
High quality sequence stop: 623.
Location/Qualifiers
1. 744

FEATURES
source
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="LD30267"
/clone_lib="LD Drosophila melanogaster embryo pot2"
/sex="male and female"
/dev_stage="0 to 24 hours mixed stage embryonic"
/lab_host="XLI Blue"
/note="Organ: embryo; Vector: pOT2; Site1: EcoRI; Site2: XhoI; Sized fractionated cDNAs were directly ligated into pOT2."
Location/Qualifiers
1. 744

BASE COUNT 186 a 238 c 197 g 123 t

ORIGIN

Query Match 78.2%; Score 17.2; DB 14; Length 744;
Best Local Similarity 86.4%; Pred. No. 3.7e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 cagccgacgtagctgcgtct 22
|||||
Db 575 CAGCCGACGGCTGCTGCTCT 596

RESULT 9
BG414942 910 bp mRNA EST 13-MAR-2001
LOCUS BG414942
DEFINITION HVSMEK0004G02f Hordeum vulgare testa/pericarp EST library
HVCNDA0013 (normal) Hordeum vulgare cDNA clone HVSMEK0004G02f, mRNA sequence.
ACCESSION BG414942
VERSION BG414942.1 GI:13320493
KEYWORDS EST.
SOURCE barley.
ORGANISM Hordeum vulgare

REFERENCE AUTHORS
Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae; Triticeae; Hordeum.
1 (bases 1 to 910)
Wing, R., Close, T.J., Kleinbols, A., Wise, R., Begum, D., Frisch, D., Yu, Y., Anderson, H., Dale, J., Henry, D., Kernodle, S., Palmer, M., Rambo, T., Saski, C., Schwartzbeck, J., Simmons, J., Choi, D.W., Main, D. and Wood, T.

TITLE
Development of a genetically and physically anchored EST resource for barley genomics

JOURNAL
Unpublished (2000)

COMMENT
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: AATTAACCTCCTCAAGG
High quality sequence stop: 799.
Location/Qualifiers

FEATURES
source
1..910
/organism="Hordeum vulgare"
/cultivar="Morex"
/db_xref="taxon:4513"
/clone="HYSMEK0004G02f"
/cdna_lib="Hordeum vulgare testa/pericarp EST library
HCDNA0013 (normal)"
/tissue_type="testa/pericarp"
/lab_host="TUC121"
/note="Vector: lambdaZAP; Site 1: EcoRI; Site 2: XhoI. For more details on library preparation and sequence analysis see <http://www.genome.clemson.edu/projects/barley/>"

BASE COUNT
191 a 243 c 289 g 184 t 3 others

ORIGIN

Query Match 78.2%; Score 17.2; DB 153; Length 910;
Best Local Similarity 86.4%; Pred. No. 3.6e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 cagccgacgtagctgcgtcct 22
|||||
Db 456 CCGCGACGCTAGCGCGCTCAT 477

RESULT 10
BE962535 958 bp mRNA EST 14-DEC-2000
LOCUS 60165584BRI NIH_MGC_66 Homo sapiens cDNA clone IMAGE:3855970 3',
DEFINITION mRNA sequence.
ACCESSION BE962535
VERSION BE962535.2 GI:11765530
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 958)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
On Oct 3, 2000 this sequence version replaced gi:10573240.
Contact: Robert Strausberg, Ph.D.
Email: cgapds-remail.nih.gov
Tissue Procurement: DCTD/DTF
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>
Plate: LHCMS70 row: 0 column: 11

FEATURES
High quality sequence stop: 101.
Location/Qualifiers
source
1..958
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3855970"
/cdna_lib="NIH MGC 66"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: ovary; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.8 kb. Library constructed by Life Technologies."

BASE COUNT
215 a 224 c 226 g 293 t

ORIGIN

Query Match 78.2%; Score 17.2; DB 142; Length 958;
Best Local Similarity 86.4%; Pred. No. 3.6e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 cagccgacgtagctgcgtcct 22
|||||
Db 235 CAGCGACGCTGCTGCGCTCT 214

RESULT 11
CNS050RJ 993 bp DNA GSS 26-MAY-2000
LOCUS CNS050RJ/C
DEFINITION Tetradon nigroviridis genome survey sequence SP6 end of clone 006015 of library B from Tetradon nigroviridis, genomic survey sequence.
ACCESSION AL347032.1 GI:8240802
VERSION AL347032.1 GI:8240802
KEYWORDS GSS: genome survey sequence.
SOURCE Tetradon nigroviridis.
ORGANISM Tetradon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetradon.
1 (bases 1 to 993)
Roest-Crollius, H., Jallion, O., Dasilva, C., Fizames, C., Fisher, C., Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and Weissenbach, J.
Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetradon nigroviridis
Unpublished
2 (bases 1 to 993)
Roest-Crollius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C., Bernot, A., Fizames, C., Wincker, P., Brothier, P., Quetier, F., Saurin, W. and Weissenbach, J.
Human gene number estimate provided by genome wide analysis using Tetradon nigroviridis DNA sequence
Unpublished
3 (bases 1 to 993)
Genoscope.
Direct Submission
Submitted (12-APR-2000) to the EMBL/Genbank/DBJ databases
This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetradon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/tetradon>.
Location/Qualifiers
source
1..993
/organism="Tetradon nigroviridis"
/db_xref="taxon:99883"
/clone="006015"
/cdna_lib="B"
/note="Genoscope sequence ID: COAB006AH08B1-end : SP6"

BASE COUNT
153 a 296 c 323 g 195 t 26 others

ORIGIN

Query Match	78.2%	Score 17.2	DB 232	Length 993
Best Local Similarity	86.4%	Pred. No. 3.6e022		
Matches 19	Conservative 0	Mismatches 3	Indels 0	Gaps 0
Qy	1	cagccgacgtagctgcgtctc	22	
Db	300	cagcagatggcagcgtgcgtct	279	

RESULT	12
AW265763	
LOCUS	AW265763 259 bp mRNA EST 28-DEC-1999
DEFINITION	x8b1e09.r.1 NCI_CGAP_Ov40 Homo sapiens cDNA clone IMAGE:2746792 3
SIMILAR TO	TR:Q92584 Q92584 MYELOBLAST KIAA0198 ; mRNA sequence
ACCESSION	AW265763
VERSION	AW265763.1 GI:6642579
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 259)	NCI-CCAP	http://www.ncbi.nlm.nih.gov/ncicgap .		
	National Cancer Institute,	Cancer Genome Anatomy Project (CGAP),		
	Tumor Gene Index			
	Unpublished (1997)			
	Contact: Robert Strausberg, Ph.D.			

JOURNAL COMMENT
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgaab5-remail.nih.gov
 Tissue Procurement: Andrew Beechuck M.D., John Gillespie M.D.,
 Michael R. Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: David B. Kitzman, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bdpr/image/image.html

Trace considered overall poor quality
Possible reversed clone: similarity on wrong strand
Seq primer: -40UP from Gibco
High quality sequence stop: 1.

FEATURES	Location/Qualifiers
source	1. .259

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2746792"
/clone_id="NCI_CGAP_Ov40"
/sex="female"
/tissue_type="endometrialoid ovarian metastasis"
/lab_host="DH10B"
/note="Organ: ovary; Vector: PAMPI0; cDNA made by oligo-dT
priming. Non-directionally cloned into the upc sites of
PAMPI0. Size-selected on agarose gel, average insert
size 500 bp. Primary library; non-amplified. cDNA
Library Preparation: David B. Krizman, Ph.D (NCI).
Reference: Krizman et al. (1996) Cancer Research
56:5380-5383."

```

Query Match	76.4%	Score 16.8	DB 113	Length 259
Best Local	Similarly	90.0%	Pred. No. 5.8e+02	
Matches 18	Conservative	0	Mismatches 2	Indels 0
				Gaps 0

QY 2 agccgacgtagctgcgtcc 21
||||| ||||| |||
Db 126 AGCCGGCGGTAGCTGGGCC 145

RESULT 13
BF825637

LOCUS	BF825637	281 bp	mRNA	EST	13-JAN-2001
DEFINITION	RC0-HN0024-161100-021-f06 HN0024 Homo sapiens CDNA, mRNA sequence				
ACCESSION	BF825637				
VERSION	BF825637.1				
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				

ORGANISM	REFERENCE	AUTHORS
Homo sapiens		
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.		
1 (bases 1 to 281)		
Dias Neto, E., Garcia Correa, R., Verjovsky-Almeida, S., Briones, M., R.		

AUTHORS
Dias Neto, E., Garcia Correa, R., Veijolahti-Almela, S., Britons, M. R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G. H., Carvalho, A.F., Matsukuma, A., Bala, G. S., Simpson, D. H., Brunstein, A., deoliveira, P. S., Bucher, P., Jongseneel, C. V., O'Hare, M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and Simpson, A. J.

TITLE	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags.
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE	20202663
COMMENT	Contact: Simpson A.J.G.

Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707003

Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?rl=RC06c2-RC0-HN0024&id=023&ref=3200011164>)

Seq primer: puc 18 forward
High quality sequence start: 54
High quality sequence stop: 281

FEATURES	Location/Qualifiers
source	1. .281

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HN0024"
/dev_stage="Adult"
/notes="Organ: head_normal; Vector: puc18; Site_1: Sma1;
Site_2: Sma1; A mini-library was made by cloning products
derived from ORNSTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

```

Query Match	76.4%	Score 16.8	DB 170	length 281
Best Local	Similarity 90.0%	Pred. No. 5.8e+02		
Matches 18	Conservative 0	Mismatches 2	Indels 0	Gaps 0

QY 2 agccgacgtagctgcctc 21
 |||| ||||| ||||
 Db 35 AGCCACACGCTAGCTGTGCC 54

RESULT	14
AU032424/c	
LOCUS	AU032424 332 bp mRNA EST 20-OCT-1998
DEFINITION	Rice root Oryza sativa cDNA clone R4067_1A, mRNA sequence.
ACCESSION	AU032424 AU032424.1 GI:3768397
VERSION	
KEYWORDS	EST.
SOURCE	Oryza sativa.
ORGANISM	Oryza sativa

ORGANISM
Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE 1 (bases 1 to 332)
 AUTHORS Minobe,Y. and Sasaki,T.
 TITLE Rice cDNA from root
 JOURNAL Unpublished (1995)
 COMMENT Contact: Takuji Sasaki
 National Institute of Agrobiological Resources
 Rice Genome Research Program
 2-1-2 Kannondai,Tsukuba
 Ibaraki,
 Japan 305
 Tel: 0298-38-7441
 Fax: 0298-38-7468
 Email: tsasaki@abr.affrc.go.jp
 PROJECT = 'RGP'

FEATURES

location/Qualifiers

1..332 /organism="Oryza sativa"
 /strain="Nipponbare, sub-species Japonica"
 /db_xref="taxon:4530"
 /clone="RA067.1A"
 /clone_1lb="Rice root"
 /note="Prepared from seedling root."

BASE COUNT 47 a 121 c 105 g 51 t 8 others
 ORIGIN

Query Match 76.4%; Score 16.8; DB 107; Length 332;
 Best Local Similarity 85.7%; Pred. No. 5.8e+02;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 cagccgacgtagctgcgtcc 21
 ||| ||||| ||| |||||
 Db 242 CAGCGACGCTGCGCGCTCC 222

RESULT 15
 BF978349 928 bp mRNA EST 22-JAN-2001
 LOCUS 602148933F2 NIH_MGC_62 Homo sapiens cDNA clone IMAGE:4307625 5',
 DEFINITION mRNA sequence.
 ACCESSION BF978349
 VERSION BF978349.1 GI:12345564
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 928)
 NIH-MGC http://mgc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapds-remail.nih.gov
 Tissue Procurement: ATCC/DCTD/DTP
 CDNA Library Preparation: CLONTECH Laboratories, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: LNCM178 row: b column: 10
 High quality sequence stop: 104.
 location/Qualifiers
 1..928

FEATURES

location/Qualifiers

1..928 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4307625"
 /clone_1lb="NIH_MGC_62"
 /tissue_type="melanotic melanoma, high MDR"
 /lab_host="DH10B (TI phage-resistant)"
 /note="Organ: skin; Vector: pDNR-LIB (Clontech); Site_1:
 Sfil (ggcgcctgcgc); Site_2: Sfil (ggccatcatgccc);
 Double-stranded cDNA was prepared from cell line RNA. 5'

and 3' adaptors were used in cloning as follows: 5'
 adaptor sequence: 5'-CACGGCATTATGCG-3' and 3' adaptor
 sequence: 5'-ATTCTAGAGCGCCGACGCGCCGCGCATG-dT(30)BN-3'
 (where B = A, C, or G and N = A, C, G, or T). Average
 insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies
 contained inserts by PCR. This library was enriched for
 full-length clones and was constructed by Clontech
 Laboratories (Palo Alto, CA)."

BASE COUNT 235 a 319 c 247 g 127 t
 ORIGIN

Query Match 74.5%; Score 16.4; DB 172; Length 928;
 Best Local Similarity 94.4%; Pred. No. 8.7e+02;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 cagccgacgtagctgcg 18
 ||| ||||| ||||| |||||
 Db 602 CACCGACGCTAGCTGCG 585

Search completed: September 20, 2001, 23:28:18
 Job time: 25558 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 21, 2001, 02:16:34 ; Search time 164.23 Seconds
(without alignments)
25.360 Million cell updates/sec

Title: US-09-138-735-9

Perfect score: 22

Sequence: 1 cagccgacgtagctgcgtcct 22

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA:*

- 1: /cgnl_7/ptodata/1/ina/5A.COMB.seq:*
- 2: /cgnl_7/ptodata/1/ina/5B.COMB.seq:*
- 3: /cgnl_7/ptodata/1/ina/6A.COMB.seq:*
- 4: /cgnl_7/ptodata/1/ina/6B.COMB.seq:*
- 5: /cgnl_7/ptodata/1/ina/PTUS.COMB.seq:*
- 6: /cgnl_7/ptodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	ID	Description
1	22	100.0	22	1 US-08-480-917-9	Sequence 9, Appli
2	22	100.0	3402	1 US-08-480-917-1	Sequence 1, Appli
3	15.6	70.9	269	4 US-08-482-918-58	Sequence 58, Appli
4	15.6	70.9	269	4 US-09-224-681-58	Sequence 58, Appli
5	15.6	70.9	269	4 US-08-336-728A-58	Sequence 58, Appli
6	15.2	69.1	139	3 US-08-471-914-8	Sequence 8, Appli
7	15.2	69.1	790	2 US-08-715-204-2	Sequence 2, Appli
8	15.2	69.1	790	2 US-09-162-597-2	Sequence 2, Appli
9	15.2	69.1	2343	2 US-09-018-760-3	Sequence 3, Appli
10	15.2	69.1	4060	1 US-08-308-949A-1	Sequence 1, Appli
11	15.2	69.1	16885	1 US-08-390-878-16	Sequence 16, Appli
12	15.2	69.1	80161	3 US-09-036-987A-1	Sequence 1, Appli
13	14.8	67.3	788	4 US-08-998-416-440	Sequence 440, App
14	14.8	67.3	1894	4 US-07-912-122-3	Sequence 3, Appli
15	14.8	67.3	1894	5 PCT-US93-06404-3	Sequence 3, Appli
16	14.8	67.3	3453	1 US-07-813-593-1	Sequence 1, Appli
17	14.8	67.3	3453	1 US-07-977-451-1	Sequence 1, Appli
18	14.8	67.3	3453	1 US-07-946-507-1	Sequence 1, Appli
19	14.8	67.3	3453	1 US-08-252-517-1	Sequence 1, Appli
20	14.8	67.3	3453	1 US-07-906-397A-1	Sequence 1, Appli
21	14.8	67.3	3453	1 US-08-601-891-1	Sequence 1, Appli
22	14.8	67.3	3453	2 US-09-021-324-1	Sequence 1, Appli
23	14.8	67.3	3453	5 PCT-US92-02750-1	Sequence 1, Appli
24	14.8	67.3	3453	5 PCT-US92-05401-1	Sequence 1, Appli
25	14.8	67.3	3453	5 PCT-US92-09893-1	Sequence 1, Appli
26	14.8	67.3	3521	1 US-08-222-299-1	Sequence 11, Appli
27	14.8	67.3	3521	2 US-08-434-878-1	Sequence 1, Appli

C 28	14.8	67.3	3521	5 PCT-US95-03718-1	Sequence 1, Appli
C 29	14.6	66.4	348	2 US-08-610-728B-3	Sequence 3, Appli
C 30	14.6	66.4	448	2 US-08-610-728B-11	Sequence 11, Appli
C 31	14.6	66.4	1207	2 US-08-525-864A-5	Sequence 5, Appli
C 32	14.6	66.4	1320	1 US-08-176-412-1	Sequence 1, Appli
C 33	14.6	66.4	1320	2 US-08-555-268A-1	Sequence 1, Appli
C 34	14.6	66.4	1320	5 US-08-495-695B-1	Sequence 1, Appli
C 35	14.6	66.4	1320	5 PCT-US94-14436-1	Sequence 1, Appli
C 36	14.6	66.4	1596	1 US-08-531-601-2	Sequence 2, Appli
C 37	14.6	66.4	1596	2 US-08-899-032-2	Sequence 2, Appli
C 38	14.6	66.4	1607	3 US-08-753-007A-3	Sequence 3, Appli
C 39	14.6	66.4	1607	3 US-09-398-496-3	Sequence 3, Appli
C 40	14.6	66.4	2467	3 US-08-753-007A-1	Sequence 1, Appli
C 41	14.6	66.4	2467	3 US-09-398-496-1	Sequence 1, Appli
C 42	14.6	66.4	3441	2 US-08-525-864A-1	Sequence 1, Appli
C 43	14.6	66.4	3825	1 US-08-737-597-1	Sequence 1, Appli
C 44	14.6	66.4	3825	1 US-08-737-597-2	Sequence 2, Appli
C 45	14.6	66.4	6312	1 US-08-531-601-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-08-480-917-9
; Sequence 9, Application US/08480917
; Patent No. 5820864
; GENERAL INFORMATION:
; APPLICANT: PARANHOS-BACCALA, Glaucia
; APPLICANT: LESENECHAL, Mylene
; APPLICANT: JOLIVET, Michel
; TITLE OF INVENTION: NEW TRYPAANOSOMA CRUZI ANTIGEN, AND GENE
; TITLE OF INVENTION: ENCODING THE LATTER, THEIR APPLICATION TO THE DETECTION OF
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Oliff & Berridge
; STREET: 700 South Washington Street, Suite 300
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480, 917
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Berridge, William P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: MPB 36400
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6400
; TELEFAX: 703-836-2787
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 22 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-480-917-9

Query Match 100.0%; Score 22; DB 1; Length 22;
Best Local Similarity 100.0%; Pred. No. 0.14;
Matches 22; Conservative 0; Mismatches 0; Gaps 0;
Oy 1 cagccgacgtagctgcgtcct 22

DB 1 CAGCCGACGTAGTGTCTCT 22

RESULT 2

US-08-480-917-1
Sequence 1, Application US/08480917
Patent No. 5820864
GENERAL INFORMATION:
APPLICANT: PARANHOS-BACCALA, Glaucaia
APPLICANT: LESENECHAL, Mylene
TITLE OF INVENTION: NEW TRYPANOSOMA CRUZI ANTIGEN, AND GENE
TITLE OF INVENTION: ENCODING THE LATTER; THEIR APPLICATION TO THE DETECTION OF
TITLE OF INVENTION: CHAGAS DISEASE
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Oliff & Berridge
STREET: 700 South Washington Street, Suite 300
CITY: Alexandria
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22314
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,917
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: MPB 36400
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ. ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3402 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-480-917-1

Query Match 100.0%; Score 22; DB 1; Length 3402;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAGCCGACGTAGTGTCTCT 22
DB 1266 CAGCCGACGTAGTGTCTCT 1287

RESULT 3
US-08-482-918-58
Sequence 58, Application US/08482918
Patent No. 6207417
GENERAL INFORMATION:
APPLICANT: Zsebo, Krisztina M.
APPLICANT: Bosseiman, Robert A.
APPLICANT: Suggs, Sidney V.
APPLICANT: Martin, Francis H.
TITLE OF INVENTION: Stem Cell Factor
NUMBER OF SEQUENCES: 104
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago

STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/482,918
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 01017/33005
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ. ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 269 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: CDS
LOCATION: join(1..210, 223..258)
US-08-482-918-58

Query Match 70.9%; Score 15.6; DB 4; Length 269;
Best Local Similarity 81.8%; Pred. No. 89;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAGCCGACGTAGTGTCTCT 22
DB 139 CTGCGCGCGTGTCTCTCT 160

RESULT 4
US-09-224-681-58
Sequence 58, Application US/09224681
Patent No. 6207454
GENERAL INFORMATION:
APPLICANT: Zsebo, Krisztina M.
APPLICANT: Bosseiman, Robert A.
APPLICANT: Suggs, Sidney V.
APPLICANT: Martin, Francis H.
TITLE OF INVENTION: Method for Enhancing the Efficiency of Gene
TITLE OF INVENTION: Transfer with Stem Cell Factor (SCF) Polypeptide
NUMBER OF SEQUENCES: 104
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/224,681
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/005,893
FILING DATE: 12-JAN-1998

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/449,653
FILING DATE: 24-MAY-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/982,255
FILING DATE: 25-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/589,701
FILING DATE: 01-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/573,616
FILING DATE: 24-AUG-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/537,198
FILING DATE: 11-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/422,383
FILING DATE: 16-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 01017/35199
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 269 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: CDS
LOCATION: join(1..210, 223..258)
US-09-224-681-58

Query Match 70.9%; Score 15.6; DB 4; Length 269;
Best Local Similarity 81.8%; Pred. No. 89;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 cagccgacgtagctgcgtcct 22
111111111111111111111111
Db 139 CTGCCGCGGTTCGCTTCCT 160
RESULT 5
US-08-336-728A-58
Sequence 58, Application US/08336728A
Patent No. 6207802
GENERAL INFORMATION:
APPLICANT: Zsebo, Krisztina M.
APPLICANT: Bosselman, Robert A.
APPLICANT: Suggs, Sidney V.
APPLICANT: Martin, Francis H.
TITLE OF INVENTION: Stem Cell Factor
NUMBER OF SEQUENCES: 104
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/336,728A
FILING DATE: 09-NOV-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/982,255
FILING DATE: 25-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/589,701
FILING DATE: 01-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/573,616
FILING DATE: 24-AUG-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/537,198
FILING DATE: 11-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/422,383
FILING DATE: 16-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 01017/32956
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 269 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: CDS
LOCATION: join(1..210, 223..258)
US-08-336-728A-58

Query Match 70.9%; Score 15.6; DB 4; Length 269;
Best Local Similarity 81.8%; Pred. No. 89;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 cagccgacgtagctgcgtcct 22
111111111111111111111111
Db 139 CTGCCGCGGTTCGCTTCCT 160
RESULT 6
US-08-471-914-8
Sequence 8, Application US/08471914A
Patent No. 6057152
GENERAL INFORMATION:
APPLICANT: Samulski, R.
APPLICANT: Xiao, X.
TITLE OF INVENTION: RECOMBINANT VIRAL VECTOR SYSTEM
FILE REFERENCE: 6636-027
CURRENT APPLICATION NUMBER: US/08/471,914A
CURRENT FILING DATE: 1995-06-06
EARLIER APPLICATION NUMBER: 08/440,738
EARLIER FILING DATE: 1995-05-15
NUMBER OF SEQ ID NOS: 13
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 8
LENGTH: 139
TYPE: DNA
ORGANISM: adeno-associated virus
US-08-471-914-8
Query Match 69.1%; Score 15.2; DB 3; Length 139;
Best Local Similarity 85.0%; Pred. No. 1.3e+02;

;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/018,760
;; FILING DATE:
;; CLASSIFICATION:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Jackson Esq., David A.
;; REGISTRATION NUMBER: 26,742
;; REFERENCE/DOCKET NUMBER: 1112-1-050
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 201-487-5800
;; TELEFAX: 201-343-1684
;; TELEX: 133521
;; INFORMATION FOR SEQ ID NO: 3:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 2343 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
;; HYPOTHETICAL: NO
;; US-09-018-760-3

Query Match 69.1%; Score 15.2; DB 2; Length 2343;
Best Local Similarity 85.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 agccgaggtagctgcgtcc 21
||||| ||||| ||||| |||||
Db 994 AGCCAGCGAGCTGCTGCC 975

RESULT 10

;; US-08-308-949A-1
;; Sequence 1, Application US/08308949A
;; Patent No. 5580703
;; GENERAL INFORMATION:
;; APPLICANT: Kolin, Robert M.
;; APPLICANT: Berns, Kenneth I.
;; APPLICANT: Linden, Ralph M.
;; TITLE OF INVENTION: Human Adeno-Associated Virus Integration
;; TITLE OF INVENTION: Site DNA and Uses Thereof
;; NUMBER OF SEQUENCES: 23
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
;; STREET: Two Militia Drive
;; CITY: Lexington
;; STATE: MA
;; COUNTRY: USA
;; ZIP: 02173
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/308,949A
;; FILING DATE: September 20, 1994
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 07/947,127
;; FILING DATE: September 27, 1992
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Carroll, Alice O.
;; REGISTRATION NUMBER: 33,542
;; REFERENCE/DOCKET NUMBER: ACC92-10F
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 617-861-9540
;; TELEFAX: 617-861-6240
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 4060 base pairs
;; TYPE: nucleic acid

;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
;; US-08-308-949A-1

Query Match 69.1%; Score 15.2; DB 1; Length 4060;
Best Local Similarity 85.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 gccgacggtagctgcctc 22
||||| ||||| ||||| |||||
Db 721 GCCGACGGATCAGCGCCCT 740

RESULT 11

;; US-08-390-878-16
;; Sequence 16, Application US/08390878
;; Patent No. 570663
;; GENERAL INFORMATION:
;; APPLICANT: Stover, Charles K.
;; APPLICANT: Mahairas, Gregory G.
;; TITLE OF INVENTION: VIRULENCE-ATTENUATING GENETIC DELETIONS
;; NUMBER OF SEQUENCES: 18
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Townsend and Townsend Kourie and Crew
;; STREET: One Market Plaza, Steuart Street Tower, 20th
;; STREET: Floor
;; CITY: San Francisco
;; STATE: California
;; COUNTRY: USA
;; ZIP: 94105

;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/390,878
;; FILING DATE: 17-FEB-1995
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Hunter, Tom
;; REGISTRATION NUMBER: 38,498
;; REFERENCE/DOCKET NUMBER: 15371A-17
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 415/543/9600
;; TELEFAX: 415/543/5043
;; INFORMATION FOR SEQ ID NO: 16:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 16885 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
;; US-08-390-878-16

Query Match 69.1%; Score 15.2; DB 1; Length 16885;
Best Local Similarity 85.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 cagccgaggtagctgcgtc 20
||||| ||||| ||||| |||||
Db 12302 CAGCCGAGCTAGCTGCC 12321

RESULT 12

;; US-09-036-987A-1
;; Sequence 1, Application US/09036987A
;; Patent No. 6143526
;; GENERAL INFORMATION:
;; APPLICANT: Baltz, Richard H.

Search completed: September 21, 2001, 02:16:39
Job time: 28232 sec

```

: SEQUENCE CHARACTERISTICS:
:   LENGTH: 1894 base pairs
:   TYPE: NUCLEIC ACID
:   STRANDEDNESS: single
:   TOPOLOGY: linear
:   MOLECULE TYPE: CDNA
:   FEATURE:
:     NAME/KEY: CDS
:     LOCATION: 31..1473
: US-07-912-122-3

```

```

Query Match          67.3%  Score 14.8; DB 4; Length 1894;
Best Local Similarity 88.9%  Pred. No. 2e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 cagccgacgtagctgcg 18
    ||||| ||||| |||||
Db 66 CAGCCGCGGTGCTGCG 49

```

```

RESULT 15
PCT-US93-06404-3/c
: Sequence 3, Application PC/TUS9306404
: GENERAL INFORMATION:
:   APPLICANT: YANG, Zhi
:   TITLE OF INVENTION: NOVEL FLK-2 AND ANALOGS
:   NUMBER OF SEQUENCES: 6
:   CORRESPONDENCE ADDRESS:
:     ADDRESSEE: Bertram I. Rowland
:     STREET: 4 Embarcadero Center, Suite 3400
:     CITY: San Francisco
:     STATE: California
:     COUNTRY: USA
:     ZIP: 94111
:   COMPUTER READABLE FORM:
:     MEDIUM TYPE: Floppy disk
:     COMPUTER: IBM PC compatible
:     OPERATING SYSTEM: PC-DOS/MS-DOS
:     SOFTWARE: Patentn Release #1.0, Version #1.25
:   CURRENT APPLICATION DATA:
:     APPLICATION NUMBER: PCT/US93/06404
:     FILING DATE: 19930707
:   CLASSIFICATION:
:     ATTORNEY/AGENT INFORMATION:
:       NAME: Rowland, Bertram I.
:       REGISTRATION NUMBER: 20,015
:       REFERENCE/DOCKET NUMBER: FP-55931/BIR
:     TELECOMMUNICATION INFORMATION:
:       TELEPHONE: (415) 789-1989
:       TELEFAX: (415) 398-3249
:     INFORMATION FOR SEQ ID NO: 3:
:       SEQUENCE CHARACTERISTICS:
:         LENGTH: 1894 base pairs
:         TYPE: nucleic acid
:         STRANDEDNESS: single
:         TOPOLOGY: linear
:         MOLECULE TYPE: CDNA
:         FEATURE:
:           NAME/KEY: CDS
:           LOCATION: 31..1473
: PCT-US93-06404-3

```

```

Query Match          67.3%  Score 14.8; DB 5; Length 1894;
Best Local Similarity 88.9%  Pred. No. 2e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 cagccgacgtagctgcg 18
    ||||| ||||| |||||
Db 66 CAGCCGCGGTGCTGCG 49

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 21, 2001, 02:19:48 : Search time 6407.28 Seconds
(without alignments)
50.738 Million cell updates/sec

Title: US-09-138-735-9

Perfect score: 1 cagccgacgtagctgcgtcct 22

Sequence:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 14155048 seqs, 7388405095 residues

Total number of hits satisfying chosen parameters: 28310096

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Pending_Patents_NA_Main:*
1: /cgnl_7/prodata/1/pna/PCRNUS.COMB.seq:*
2: /cgnl_7/prodata/1/pna/US06.COMB.seq:*
3: /cgnl_7/prodata/1/pna/US07.COMB.seq:*
4: /cgnl_7/prodata/1/pna/US08.COMB.seq:*
5: /cgnl_7/prodata/1/pna/US081.COMB.seq:*
6: /cgnl_7/prodata/1/pna/US082.COMB.seq:*
7: /cgnl_7/prodata/1/pna/US083.COMB.seq:*
8: /cgnl_7/prodata/1/pna/US084.COMB.seq:*
9: /cgnl_7/prodata/1/pna/US085.COMB.seq:*
10: /cgnl_7/prodata/1/pna/US086.COMB.seq:*
11: /cgnl_7/prodata/1/pna/US087.COMB.seq:*
12: /cgnl_7/prodata/1/pna/US088.COMB.seq:*
13: /cgnl_7/prodata/1/pna/US089.COMB.seq:*
14: /cgnl_7/prodata/1/pna/US090.COMB.seq:*
15: /cgnl_7/prodata/1/pna/US091.COMB.seq:*
16: /cgnl_7/prodata/1/pna/US092.COMB.seq:*
17: /cgnl_7/prodata/1/pna/US093.COMB.seq:*
18: /cgnl_7/prodata/1/pna/US094.COMB.seq:*
19: /cgnl_7/prodata/1/pna/US095A.COMB.seq:*
20: /cgnl_7/prodata/1/pna/US095B.COMB.seq:*
21: /cgnl_7/prodata/1/pna/US095C.COMB.seq:*
22: /cgnl_7/prodata/1/pna/US095D.COMB.seq:*
23: /cgnl_7/prodata/1/pna/US096A.COMB.seq:*
24: /cgnl_7/prodata/1/pna/US096B.COMB.seq:*
25: /cgnl_7/prodata/1/pna/US096C.COMB.seq:*
26: /cgnl_7/prodata/1/pna/US096D.COMB.seq:*
27: /cgnl_7/prodata/1/pna/US096E.COMB.seq:*
28: /cgnl_7/prodata/1/pna/US097A.COMB.seq:*
29: /cgnl_7/prodata/1/pna/US097B.COMB.seq:*
30: /cgnl_7/prodata/1/pna/US097C.COMB.seq:*
31: /cgnl_7/prodata/1/pna/US098.COMB.seq:*
32: /cgnl_7/prodata/1/pna/US6000.COMB.seq:*
33: /cgnl_7/prodata/1/pna/US6001.COMB.seq:*
34: /cgnl_7/prodata/1/pna/US6002.COMB.seq:*
35: /cgnl_7/prodata/1/pna/US6003.COMB.seq:*
36: /cgnl_7/prodata/1/pna/US6004.COMB.seq:*
37: /cgnl_7/prodata/1/pna/US6005.COMB.seq:*
38: /cgnl_7/prodata/1/pna/US6006.COMB.seq:*
39: /cgnl_7/prodata/1/pna/US6007.COMB.seq:*
40: /cgnl_7/prodata/1/pna/US6008.COMB.seq:*
41: /cgnl_7/prodata/1/pna/US6009.COMB.seq:*
42: /cgnl_7/prodata/1/pna/US6010.COMB.seq:*
43: /cgnl_7/prodata/1/pna/US6011.COMB.seq:*

44: /cgnl_7/prodata/1/pna/US6012.COMB.seq:*
45: /cgnl_7/prodata/1/pna/US6013.COMB.seq:*
46: /cgnl_7/prodata/1/pna/US6014.COMB.seq:*
47: /cgnl_7/prodata/1/pna/US6015.COMB.seq:*
48: /cgnl_7/prodata/1/pna/US6016.COMB.seq:*
49: /cgnl_7/prodata/1/pna/US6017.COMB.seq:*
50: /cgnl_7/prodata/1/pna/US6018.COMB.seq:*
51: /cgnl_7/prodata/1/pna/US6019.COMB.seq:*
52: /cgnl_7/prodata/1/pna/US6020.COMB.seq:*
53: /cgnl_7/prodata/1/pna/US6021.COMB.seq:*
54: /cgnl_7/prodata/1/pna/US6022.COMB.seq:*
55: /cgnl_7/prodata/1/pna/US6023.COMB.seq:*
56: /cgnl_7/prodata/1/pna/US6024.COMB.seq:*
57: /cgnl_7/prodata/1/pna/US6025.COMB.seq:*
58: /cgnl_7/prodata/1/pna/US6026.COMB.seq:*
59: /cgnl_7/prodata/1/pna/US6027.COMB.seq:*
60: /cgnl_7/prodata/1/pna/US6028.COMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	22	100.0	22	13 US-08-988-242-5	Sequence 5, App1
2	22	100.0	22	15 US-09-138-736-9	Sequence 9, App1
3	22	100.0	3402	13 US-08-988-242-1	Sequence 1, App1
4	22	100.0	3402	15 US-09-138-736-1	Sequence 1, App1
5	17.8	80.9	465	16 US-09-270-8498-93552	Sequence 9355, A
6	17.2	78.2	87	16 US-09-270-8498-46264	Sequence 4626, A
7	17.2	78.2	253	16 US-09-244-000A-76417	Sequence 76417, A
8	17.2	78.2	264	15 US-09-654-617-106410	Sequence 106410, A
9	17.2	78.2	264	27 US-09-684-016-106410	Sequence 106410, A
10	17.2	78.2	425	25 US-09-654-617-109675	Sequence 109675, A
11	17.2	78.2	425	27 US-09-684-016-109675	Sequence 109675, A
12	17.2	78.2	426	16 US-09-270-8498-122012	Sequence 122012, A
13	17.2	78.2	2339	49 US-60-173-464-23049	Sequence 23049, A
14	17.2	78.2	2339	51 US-60-191-637-28973	Sequence 28973, A
15	17.2	78.2	2339	51 US-60-191-681-23456	Sequence 23456, A
16	17.2	78.2	4133	24 US-09-620-392-11665	Sequence 11665, A
17	17.2	78.2	7215	24 US-09-620-392-9977	Sequence 9977, A
18	17.2	78.2	9348	24 US-09-620-392-25910	Sequence 25910, A
19	17.2	78.2	9520	24 US-09-620-392-30257	Sequence 30257, A
20	17.2	78.2	9848	28 US-09-702-134-5934	Sequence 5934, A
21	17.2	78.2	10869	24 US-09-620-392-568	Sequence 568, A
22	17.2	78.2	12654	28 US-09-620-392-32904	Sequence 32904, A
23	17.2	78.2	13891	24 US-09-702-134-25170	Sequence 25170, A
24	17.2	78.2	14396	24 US-09-620-392-65700	Sequence 65700, A
25	17.2	78.2	14396	24 US-09-620-392-58244	Sequence 58244, A
26	17.2	78.2	19687	24 US-09-620-392-13752	Sequence 13752, A
27	17.2	78.2	27082	1 PCT-US01-01354-25259	Sequence 25259, A
28	17.2	78.2	29326	24 US-09-620-392-48151	Sequence 48151, A
29	17.2	78.2	29595	49 US-60-171-627-625	Sequence 625, A
30	17.2	78.2	29595	49 US-60-173-464-6116	Sequence 6116, A
31	17.2	78.2	30905	28 US-09-702-134-7927	Sequence 7927, A
32	17.2	78.2	30978	19 US-09-528-237A-1201	Sequence 1201, A
33	17.2	78.2	34222	24 US-09-620-392-26868	Sequence 26868, A
34	17.2	78.2	35362	24 US-09-620-392-44557	Sequence 44557, A
35	17.2	78.2	38678	28 US-09-702-134-5608	Sequence 5608, A
36	17.2	78.2	38699	28 US-09-620-392-13645	Sequence 13645, A
37	17.2	78.2	57538	24 US-09-702-134-6237	Sequence 6237, A
38	16.8	76.4	402	18 US-09-436-762A-27139	Sequence 27139, A
39	16.8	76.4	527	16 US-09-270-8498-151280	Sequence 151280, A
40	16.8	76.4	2540	52 US-60-207-216-209	Sequence 209, A
41	16.8	76.4	66915	48 US-60-161-932-304	Sequence 304, A
42	16.4	74.5	417	16 US-09-270-8498-56976	Sequence 56976, A
43	16.4	74.5	478	16 US-09-270-8498-121200	Sequence 121200, A
44	16.4	74.5	498	16 US-09-270-8498-39425	Sequence 39425, A
45	16.4	74.5	107600	20 US-09-534-859-571	Sequence 571, A

ALIGNMENTS

RESULT 1
US-08-988-242-5
Sequence 5, Application US/08988242
GENERAL INFORMATION:
APPLICANT: PARANHOS-BACCALA, GLAUCIA
APPLICANT: LESENECHAL, MYLENE
APPLICANT: JOLIVET, MICHEL
TITLE OF INVENTION: NEW TRYPANOSOMA CRUZI ANTIGEN, GENE
TITLE OF INVENTION: ENCODING THEREFOR, AND METHODS OF DETECTING AND TREATING
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESSES:
ADDRESSEE: OLIFF & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: Alexandria
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/988,242
FILING DATE: 10-DEC-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 36400A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 22 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-988-242-5

Query Match 100.0%; Score 22; DB 13; Length 22;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cagccgacgtagctgcctcct 22
|||||
DB 1 CAGCCGACGCTAGCTGCTCCT 22

RESULT 2
US-09-138-736-9
Sequence 9, Application US/09138736
GENERAL INFORMATION:
APPLICANT: PARANHOS-BACCALA, GLAUCIA
APPLICANT: LESENECHAL, MYLENE
APPLICANT: JOLIVET, MICHEL
TITLE OF INVENTION: NEW TRYPANOSOMA CRUZI ANTIGEN, AND GENE
TITLE OF INVENTION: ENCODING THE LATTER; THEIR APPLICATION TO THE DETECTION OF
TITLE OF INVENTION: CHAGAS DISEASE
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Oliff & Berridge
STREET: 700 South Washington Street, Suite 300
CITY: Alexandria

STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22314
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/138,736
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/480,917
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 36400
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 22 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-138-736-9

Query Match 100.0%; Score 22; DB 15; Length 22;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cagccgacgtagctgcctcct 22
|||||
DB 1 CAGCCGACGCTAGCTGCTCCT 22

RESULT 3
US-08-988-242-1
Sequence 1, Application US/08988242
GENERAL INFORMATION:
APPLICANT: PARANHOS-BACCALA, GLAUCIA
APPLICANT: LESENECHAL, MYLENE
APPLICANT: JOLIVET, MICHEL
TITLE OF INVENTION: NEW TRYPANOSOMA CRUZI ANTIGEN, GENE
TITLE OF INVENTION: ENCODING THEREFOR, AND METHODS OF DETECTING AND TREATING
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESSES:
ADDRESSEE: OLIFF & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: Alexandria
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/988,242
FILING DATE: 10-DEC-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 36400A

TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3402 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-988-242-1

Query Match 100.0%; Score 22; DB 13; Length 3402;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cagccgacgtagctgcgtcct 22
|||||
Db 1266 CAGCCGACGCTAGCTGCTCCT 1287

RESULT 4
US-09-138-736-1
Sequence 1, Application US/09138736
GENERAL INFORMATION:
APPLICANT: PARANHOS-BACCALA, Glaucia
APPLICANT: LESNECHAL, Mylene
APPLICANT: JOLIVET, Michel
TITLE OF INVENTION: NEW TRYPAOSOMA CRUZI ANTIGEN, AND GENE
TITLE OF INVENTION: ENCODING THE LATTER; THEIR APPLICATION TO THE DETECTION OF
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Oliff & Berridge
STREET: 700 South Washington Street, Suite 300
CITY: Alexandria
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22314
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/138,736
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/480,917
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 36400
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3402 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-09-138-736-1

Query Match 100.0%; Score 22; DB 15; Length 3402;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cagccgacgtagctgcgtcct 22
|||||
Db 1266 CAGCCGACGCTAGCTGCTCCT 1287

RESULT 5
US-09-270-849B-93552/C
Sequence 93552, Application US/09270849B
GENERAL INFORMATION:
APPLICANT: Swimmer et al.
TITLE OF INVENTION: Insect genome survey devices
FILE REFERENCE:
CURRENT APPLICATION NUMBER: US/09/270,849B
CURRENT FILING DATE: 1999-03-17
NUMBER OF SEQ ID NOS: 195450
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 93552
LENGTH: 465
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-270-849B-93552

Query Match 80.9%; Score 17.8; DB 16; Length 465;
Best Local Similarity 90.5%; Pred. No. 3.1e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 acccgacgtagctgcgtcct 22
|||||
Db 375 ACCCGACGCTAGCTGCTTCT 355

RESULT 6
US-09-270-849B-46264/C
Sequence 46264, Application US/09270849B
GENERAL INFORMATION:
APPLICANT: Swimmer et al.
TITLE OF INVENTION: Insect genome survey devices
FILE REFERENCE:
CURRENT APPLICATION NUMBER: US/09/270,849B
CURRENT FILING DATE: 1999-03-17
NUMBER OF SEQ ID NOS: 195450
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 46264
LENGTH: 87
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-270-849B-46264

Query Match 78.2%; Score 17.2; DB 16; Length 87;
Best Local Similarity 86.4%; Pred. No. 5.8e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 cagccgacgtagctgcgtcct 22
|||||
Db 33 CAGCCGACGCTAGCTGCTTCT 12

RESULT 7
US-09-244-000A-76417
Sequence 76417, Application US/09244000A
GENERAL INFORMATION:
APPLICANT: Byrum, Joseph R.
APPLICANT: La Rosa, Thomas J.
APPLICANT: Thompson, Michael D.
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated w/1
FILE REFERENCE: 38-21(15305)B

```
; CURRENT APPLICATION NUMBER: US/09/244,000A
; CURRENT FILING DATE: 1999-02-08
; PRIOR APPLICATION DATA removed - refer to PALM or File Wrapper
; NUMBER OF SEQ ID NOS: 101193
; SEQ ID NO 76417
; LENGTH: 253
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: 701010921H1
US-09-244-000A-76417

Query Match      78.2%; Score 17.2; DB 16; Length 253;
Best Local Similarity 86.4%; Pred. No. 5.8e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 cagccgacggtagctgcctct 22
        ||||||| ||||| ||
DB      84 cagccgacggtgctgcctct 105

RESULT      8
US-09-654-617-106410
; Sequence 106410, Application US/09654617
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; TITLE OF INVENTION: Annotated Plant Genes
; FILE REFERENCE: 38-21(15097)D
; CURRENT APPLICATION NUMBER: US/09/654,617
; CURRENT FILING DATE: 2000-09-05
; NUMBER OF SEQ ID NOS: 463173
; SEQ ID NO 106410
; LENGTH: 264
; TYPE: DNA
; ORGANISM: Glycine max
US-09-654-617-106410

Query Match      78.2%; Score 17.2; DB 25; Length 264;
Best Local Similarity 86.4%; Pred. No. 5.8e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 cagccgacggtagctgcctct 22
        ||||||| ||||| ||
DB      85 cagccgacggtgctgcctct 106

RESULT      9
US-09-684-016-106410
; Sequence 106410, Application US/09684016
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; TITLE OF INVENTION: Annotated Plant Genes
; FILE REFERENCE: 38-21(15097)D
; CURRENT APPLICATION NUMBER: US/09/684,016
; CURRENT FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: US 09/654,617
; PRIOR FILING DATE: 2000-09-05
; NUMBER OF SEQ ID NOS: 463173
; SEQ ID NO 106410
; LENGTH: 264
; TYPE: DNA
; ORGANISM: Glycine max
US-09-684-016-106410

Query Match      78.2%; Score 17.2; DB 27; Length 264;
Best Local Similarity 86.4%; Pred. No. 5.8e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 cagccgacggtagctgcctct 22
        ||||||| ||||| ||
DB      85 cagccgacggtgctgcctct 106
```

```
||||||| ||||| ||
DB      85 cagccgacggtgctgcctct 106

RESULT      10
US-09-654-617-109675
; Sequence 109675, Application US/09654617
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; TITLE OF INVENTION: Annotated Plant Genes
; FILE REFERENCE: 38-21(15097)D
; CURRENT APPLICATION NUMBER: US/09/654,617
; CURRENT FILING DATE: 2000-09-05
; NUMBER OF SEQ ID NOS: 463173
; SEQ ID NO 109675
; LENGTH: 425
; TYPE: DNA
; ORGANISM: Glycine max
US-09-654-617-109675

Query Match      78.2%; Score 17.2; DB 25; Length 425;
Best Local Similarity 86.4%; Pred. No. 5.9e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 cagccgacggtagctgcctct 22
        ||||||| ||||| ||
DB      227 cagccgacggtgctgcctct 248

RESULT      11
US-09-684-016-109675
; Sequence 109675, Application US/09684016
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; TITLE OF INVENTION: Annotated Plant Genes
; FILE REFERENCE: 38-21(15097)D
; CURRENT APPLICATION NUMBER: US/09/684,016
; CURRENT FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: US 09/654,617
; PRIOR FILING DATE: 2000-09-05
; NUMBER OF SEQ ID NOS: 463173
; SEQ ID NO 109675
; LENGTH: 425
; TYPE: DNA
; ORGANISM: Glycine max
US-09-684-016-109675

Query Match      78.2%; Score 17.2; DB 27; Length 425;
Best Local Similarity 86.4%; Pred. No. 5.9e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 cagccgacggtagctgcctct 22
        ||||||| ||||| ||
DB      227 cagccgacggtgctgcctct 248

RESULT      12
US-09-270-849B-122012/c
; Sequence 122012, Application US/09270849B
; GENERAL INFORMATION:
; APPLICANT: Swimmer et al.
; TITLE OF INVENTION: Insect genome survey devices
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/09/270,849B
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 195450
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 122012
; LENGTH: 426
```

```

; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-270-849B-122012

```

```

Query Match          78.2%; Score 17.2; DB 16; Length 426;
Best Local Similarity 86.4%; Pred. No. 5.9e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

OY 1 cagccgacgtagctgcgtcct 22
    ||||| ||||| |||||
Db 247 CAGCCGCCGGTAGCAGTGTCTT 226

```

```

RESULT 13
US-60-173-464-23049/c
; Sequence 23049, Application US/60173464
; GENERAL INFORMATION:
; APPLICANT: Li, Peter W.D.
; TITLE OF INVENTION: ISOLATED G-PROTEIN COUPLED RECEPTORS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING GPCR PROTEINS AND USES
; FILE REFERENCE: C1000173
; CURRENT APPLICATION NUMBER: US/60/173,464
; CURRENT FILING DATE: 1999-12-29
; NUMBER OF SEQ. ID NOS: 30269
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23049
; LENGTH: 2339
; TYPE: DNA
; ORGANISM: Drosophila
US-60-173-464-23049

```

```

Query Match          78.2%; Score 17.2; DB 49; Length 2339;
Best Local Similarity 86.4%; Pred. No. 6e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

OY 1 cagccgacgtagctgcgtcct 22
    ||||| ||||| |||||
Db 654 CAGCTGACCGGAGCTGCGTCTT 633

```

```

RESULT 14
US-60-191-637-28973/c
; Sequence 28973, Application US/60191637
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; TITLE OF INVENTION: NUCLEIC ACID DETECTION KITS COMPRISING
; TITLE OF INVENTION: GENE SEQUENCES EXPRESSED FROM THE DROSOPHILA GENOME, AND
; FILE REFERENCE: C1000392
; CURRENT APPLICATION NUMBER: US/60/191,637
; CURRENT FILING DATE: 2000-03-23
; NUMBER OF SEQ. ID NOS: 42660
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28973
; LENGTH: 2339
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-60-191-637-28973

```

```

Query Match          78.2%; Score 17.2; DB 51; Length 2339;
Best Local Similarity 86.4%; Pred. No. 6e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

OY 1 cagccgacgtagctgcgtcct 22
    ||||| ||||| |||||
Db 654 CAGCTGACCGGAGCTGCGTCTT 633

```

```

RESULT 15
US-60-191-681-23456/c
; Sequence 23456, Application US/60191681
; GENERAL INFORMATION:
; APPLICANT: Li, Peter, W.D.
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING ISOLATED DROSOPHILA PROTEINS AND
; FILE REFERENCE: C1000390
; CURRENT APPLICATION NUMBER: US/60/191,681
; CURRENT FILING DATE: 2000-03-23
; NUMBER OF SEQ. ID NOS: 30973
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23456
; LENGTH: 2339
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-60-191-681-23456

```

```

Query Match          78.2%; Score 17.2; DB 51; Length 2339;
Best Local Similarity 86.4%; Pred. No. 6e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

OY 1 cagccgacgtagctgcgtcct 22
    ||||| ||||| |||||
Db 654 CAGCTGACCGGAGCTGCGTCTT 633

```

```

Search completed: September 21, 2001, 02:19:54
Job time: 30466 sec

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 21, 2001, 02:39:19 ; Search time 1035.34 Seconds
(without alignments)
42.408 Million cell updates/sec

Title: US-09-138-735-9

Perfect score: 22

Sequence: 1 cagccgacgtagctgcgtcct 22

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1741792 seqs, 997871973 residues

Total number of hits satisfying chosen parameters: 3483584

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Pending_Patents_NA_New:*

1: /cgnl_7/ptodata/2/pna/PCY_NEW_COMB.seq:*\n2: /cgnl_7/ptodata/2/pna/US06_NEW_COMB.seq:*\n3: /cgnl_7/ptodata/2/pna/US07_NEW_COMB.seq:*\n4: /cgnl_7/ptodata/2/pna/US08_NEW_COMB.seq:*\n5: /cgnl_7/ptodata/2/pna/US09_NEW_COMB.seq:*\n6: /cgnl_7/ptodata/2/pna/US09_NEW_COMB.seq1:*\n7: /cgnl_7/ptodata/2/pna/US09_NEW_COMB.seq2:*\n8: /cgnl_7/ptodata/2/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	22	100.0	22	US-09-138-735-9	Sequence 9, Appli
2	22	100.0	3402	US-09-138-735-1	Sequence 1, Appli
3	17.2	78.2	27082	US-09-764-905-25259	Sequence 25259, A
4	16.4	74.5	107600	US-09-803-736-571	Sequence 571, App
5	16.2	73.6	194	US-60-252-833-32177	Sequence 32177, A
6	16.2	73.6	226	US-60-252-833-25761	Sequence 25761, A
7	16.2	73.6	264	US-09-864-761-24156	Sequence 24156, A
8	16.2	73.6	272	US-60-252-833-25819	Sequence 25819, A
9	16.2	73.6	273	US-09-912-292-55206	Sequence 55206, A
10	16.2	73.6	277	US-60-252-833-2481	Sequence 2481, Ap
11	16.2	73.6	326	US-60-252-833-32199	Sequence 32199, A
12	16.2	73.6	327	US-60-252-833-13183	Sequence 13183, A
13	16.2	73.6	332	US-09-540-213-15508	Sequence 15508, A
14	16.2	73.6	345	US-60-252-833-14207	Sequence 14207, A
15	16.2	73.6	369	US-60-252-833-19140	Sequence 19140, A
16	16.2	73.6	378	US-09-909-627-8996	Sequence 8996, Ap
17	16.2	73.6	380	US-60-252-833-32484	Sequence 32484, A
18	16.2	73.6	412	US-60-252-833-32592	Sequence 32592, A
19	16.2	73.6	412	US-60-252-833-32600	Sequence 32600, A
20	16.2	73.6	421	US-60-253-651-22049	Sequence 22049, A
21	16.2	73.6	471	US-09-866-555-17442	Sequence 17442, A
22	16.2	73.6	482	US-60-253-653-3881	Sequence 3881, Ap
23	16.2	73.6	491	US-09-849-626-1155	Sequence 1155, Ap
24	16.2	73.6	491	US-09-902-941-1155	Sequence 1155, Ap
25	16.2	73.6	493	US-09-834-366-50093	Sequence 50093, A

26	16.2	73.6	520	US-09-849-626-927	Sequence 927, App
27	16.2	73.6	520	US-09-902-941-927	Sequence 927, App
28	16.2	73.6	540	US-60-253-653-15892	Sequence 15892, A
29	16.2	73.6	542	US-60-252-833-30020	Sequence 30020, A
30	16.2	73.6	546	US-60-253-652-16860	Sequence 16860, A
31	16.2	73.6	593	US-09-864-761-7514	Sequence 7514, Ap
32	16.2	73.6	613	US-60-253-654-10376	Sequence 10376, A
33	16.2	73.6	613	US-60-255-592-10376	Sequence 10376, A
34	16.2	73.6	725	US-60-252-833-9958	Sequence 9958, Ap
35	16.2	73.6	731	US-60-253-652-14151	Sequence 14151, A
36	16.2	73.6	740	US-60-252-833-27687	Sequence 27687, A
37	16.2	73.6	1029	US-60-312-544-1195	Sequence 1195, Ap
38	16.2	73.6	5642	US-60-278-561-9583	Sequence 9583, Ap
39	16.2	73.6	5643	US-09-898-888-11040	Sequence 11040, A
40	15.8	71.8	298	US-60-253-654-18217	Sequence 18217, A
41	15.8	71.8	298	US-60-255-592-18217	Sequence 18217, A
42	15.8	71.8	383	US-60-253-456-3926	Sequence 3926, Ap
43	15.8	71.8	404	US-09-849-626-443	Sequence 443, App
44	15.8	71.8	404	US-09-902-941-443	Sequence 443, App
45	15.8	71.8	5031	US-09-764-877-2950	Sequence 2950, Ap

ALIGNMENTS

```

RESULT 1
US-09-138-735-9
; Sequence 9, Application US/09138735
; GENERAL INFORMATION:
; APPLICANT: PARANHOS-BACCALA, Glaucia
; APPLICANT: LESENECHAL, Mylene
; APPLICANT: JOLIVER, Michel
; TITLE OF INVENTION: TYPANOSOMA CRUZI ANTIGEN, GENE ENCODING THEREFOR AND METHODS
; FILE REFERENCE: WPB 36400B
; CURRENT APPLICATION NUMBER: US/09/138,735
; PRIOR FILING DATE: 1998-08-24
; PRIOR APPLICATION NUMBER: US 08/480,917
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: FR 94/10132
; PRIOR FILING DATE: 1994-08-12
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
; LENGTH: 22
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: probe/primer
US-09-138-735-9

Query Match 100.0%; Score 22; DB 6; Length 22;
Best Local Similarity 100.0%; Pred. No. 0.26;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cagccgacgtagctgcgtcct 22
Db 1 cagccgacgtagctgcgtcct 22

RESULT 2
US-09-138-735-1
; Sequence 1, Application US/09138735
; GENERAL INFORMATION:
; APPLICANT: PARANHOS-BACCALA, Glaucia
; APPLICANT: LESENECHAL, Mylene
; APPLICANT: JOLIVER, Michel
; TITLE OF INVENTION: TYPANOSOMA CRUZI ANTIGEN, GENE ENCODING THEREFOR AND METHODS
; FILE REFERENCE: WPB 36400B
; CURRENT APPLICATION NUMBER: US/09/138,735
; CURRENT FILING DATE: 1998-08-24

```

1 PRIOR APPLICATION NUMBER: US 08/480,917
PRIOR FILING DATE: 1995-06-07
PRIOR APPLICATION NUMBER: FR 94/10132
PRIOR FILING DATE: 1994-08-12
NUMBER OF SEQ ID NOS: 13
SOFTWARE: PatentIn version 3.0
SEQ ID NO 1
LENGTH: 3402
TYPE: DNA
ORGANISM: Trypanosoma cruzi
US-09-138-735-1

Query Match 100.0%; Score 22; DB 6; Length 3402;
Best Local Similarity 100.0%; Pred. No. 0.29;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cagcgacggtacgtcgtcct 22
1266 cagcgacggtacgtcgtcct 1287

RESULT 3
US-09-764-905-25259
Sequence 25259, Application US/09764905
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC004
CURRENT APPLICATION NUMBER: US/09/764,905
CURRENT FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: 60/179,065
PRIOR FILING DATE: 2000-01-31
PRIOR APPLICATION NUMBER: 60/180,628
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: 60/214,886
PRIOR FILING DATE: 2000-06-28
PRIOR APPLICATION NUMBER: 60/217,487
PRIOR FILING DATE: 2000-07-11
PRIOR APPLICATION NUMBER: 60/225,758
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/220,963
PRIOR FILING DATE: 2000-07-26
PRIOR APPLICATION NUMBER: 60/217,496
PRIOR FILING DATE: 2000-07-11
PRIOR APPLICATION NUMBER: 60/225,447
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/218,290
PRIOR FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: 60/225,757
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/226,868
PRIOR FILING DATE: 2000-08-22
PRIOR APPLICATION NUMBER: 60/216,647
PRIOR FILING DATE: 2000-07-07
PRIOR APPLICATION NUMBER: 60/225,267
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/216,880
PRIOR FILING DATE: 2000-07-07
PRIOR APPLICATION NUMBER: 60/225,270
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/251,869
PRIOR FILING DATE: 2000-12-08
PRIOR APPLICATION NUMBER: 60/235,834
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: 60/234,274
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: 60/234,223
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: 60/228,924
PRIOR FILING DATE: 2000-08-30
PRIOR APPLICATION NUMBER: 60/224,518
PRIOR FILING DATE: 2000-08-14

PRIOR APPLICATION NUMBER: 60/236,369
PRIOR FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: 60/224,519
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/220,964
PRIOR FILING DATE: 2000-07-26
PRIOR APPLICATION NUMBER: 60/241,809
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/249,299
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/236,327
PRIOR FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: 60/241,785
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/244,617
PRIOR FILING DATE: 2000-11-01
PRIOR APPLICATION NUMBER: 60/225,268
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/236,368
PRIOR FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: 60/251,856
PRIOR FILING DATE: 2000-12-08
PRIOR APPLICATION NUMBER: 60/251,868
PRIOR FILING DATE: 2000-12-08
PRIOR APPLICATION NUMBER: 60/229,344
PRIOR FILING DATE: 2000-09-01
PRIOR APPLICATION NUMBER: 60/234,997
PRIOR FILING DATE: 2000-09-25
PRIOR APPLICATION NUMBER: 60/229,343
PRIOR FILING DATE: 2000-09-01
PRIOR APPLICATION NUMBER: 60/229,345
PRIOR FILING DATE: 2000-09-01
PRIOR APPLICATION NUMBER: 60/229,287
PRIOR FILING DATE: 2000-09-01
PRIOR APPLICATION NUMBER: 60/229,513
PRIOR FILING DATE: 2000-09-05
PRIOR APPLICATION NUMBER: 60/231,413
PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: 60/229,509
PRIOR FILING DATE: 2000-09-05
PRIOR APPLICATION NUMBER: 60/236,367
PRIOR FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: 60/237,039
PRIOR FILING DATE: 2000-10-02
PRIOR APPLICATION NUMBER: 60/237,038
PRIOR FILING DATE: 2000-10-02
PRIOR APPLICATION NUMBER: 60/236,370
PRIOR FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: 60/236,802
PRIOR FILING DATE: 2000-10-02
PRIOR APPLICATION NUMBER: 60/237,037
PRIOR FILING DATE: 2000-10-02
PRIOR APPLICATION NUMBER: 60/237,040
PRIOR FILING DATE: 2000-10-02
PRIOR APPLICATION NUMBER: 60/240,960
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/239,935
PRIOR FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: 60/239,937
PRIOR FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: 60/241,787
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/246,474
PRIOR FILING DATE: 2000-11-08
PRIOR APPLICATION NUMBER: 60/246,532
PRIOR FILING DATE: 2000-11-08
PRIOR APPLICATION NUMBER: 60/249,216
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,210
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/226,681
PRIOR FILING DATE: 2000-08-22
PRIOR APPLICATION NUMBER: 60/225,759

```

; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/225,213
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/227,182
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/225,214
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/235,836
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/230,438
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/215,135
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 60/225,266
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/249,218
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,208
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,213
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,212
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,207
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,245
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,244
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,217
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,211
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,215
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,264
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,214
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,297
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/232,400
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/231,242
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/232,081
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/232,080
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,414
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,244
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/233,064
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/233,063
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,397
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,399
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,401
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/241,808
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,826
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,786
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,221
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/246,475
; PRIOR FILING DATE: 2000-11-08

; PRIOR APPLICATION NUMBER: 60/231,243
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/233,065
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,398

Query Match      78.2%; Score 17.2; DB 7; Length 27082;
Best Local Similarity 86.4%; Pred. No. 63;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1  cagccgacgtagctgcgtcct 22
        ||||| ||||| ||||| |||||
Db      531  cagcccgacgtagctgcgtcct 552

RESULT      4
US-09-803-736-571/c
; Sequence 571, Application US/09803736
; GENERAL INFORMATION:
; APPLICANT: Bush, David F.
; APPLICANT: Levin, Irene M.
; APPLICANT: Norris, Susan R.
; APPLICANT: Rounsley, Steven D.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Plant Polymorphic Markers and Uses Thereof
; FILE REFERENCE: 38-10(15493)D
; CURRENT APPLICATION NUMBER: US/09/803,736
; CURRENT FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 09/534,859
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: Identified by Attorney docket number 04983.0206CFSU01 38
; PRIOR FILING DATE: 2000-10-20
; NUMBER OF SEQ ID NOS: 1582
; SEQ ID NO 571
; LENGTH: 107600
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-803-736-571

Query Match      74.5%; Score 16.4; DB 6; Length 107600;
Best Local Similarity 94.4%; Pred. No. 1.6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      5  cgaagtagctgcgtcct 22
        ||||| ||||| ||||| |||||
Db      33860  CGACGTAAGCTGCTTCT 33843

RESULT      5
US-60-252-833-32177/c
; Sequence 32177, Application US/60252833
; GENERAL INFORMATION:
; APPLICANT: Glenn, Matthew
; APPLICANT: Grigor, Murray R.
; TITLE OF INVENTION: Compositions isolated from bovine
; TITLE OF INVENTION: Clissues and methods for their use.
; FILE REFERENCE: 1052P2
; CURRENT APPLICATION NUMBER: US/60/252,833
; CURRENT FILING DATE: 2000-11-21
; NUMBER OF SEQ ID NOS: 43535
; SOFTWARE: PASTESQ for Windows Version 4.0
; SEQ ID NO 32177
; LENGTH: 194
; TYPE: DNA
; ORGANISM: Bovine
US-60-252-833-32177

Query Match      73.6%; Score 16.2; DB 8; Length 194;
Best Local Similarity 85.7%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```


GENERAL INFORMATION:
APPLICANT: Rosen, et. al.
TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 101
FILE REFERENCE: PO-101
CURRENT FILING DATE: 2001-07-26
PRIOR APPLICATION NUMBER: 08/103,746
PRIOR FILING DATE: 1993-08-09
PRIOR APPLICATION NUMBER: 09/859,417
PRIOR FILING DATE: 2001-05-18
PRIOR APPLICATION NUMBER: 08/103,743
PRIOR FILING DATE: 1993-08-09
PRIOR APPLICATION NUMBER: 09/741,827
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 09/813,154
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 08/196,482
PRIOR FILING DATE: 1994-02-15
PRIOR APPLICATION NUMBER: 09/783,587
PRIOR FILING DATE: 2001-02-15
PRIOR APPLICATION NUMBER: 08/196,481
PRIOR FILING DATE: 1994-02-15
PRIOR APPLICATION NUMBER: 08/220,662
PRIOR FILING DATE: 1994-03-31
PRIOR APPLICATION NUMBER: 08/220,661
PRIOR FILING DATE: 1994-03-31
PRIOR APPLICATION NUMBER: 08/275,627
PRIOR FILING DATE: 1994-07-15
PRIOR APPLICATION NUMBER: 08/345,704
PRIOR FILING DATE: 1994-11-21
PRIOR APPLICATION NUMBER: 09/859,662
PRIOR FILING DATE: 2001-05-18
PRIOR APPLICATION NUMBER: 08/401,881
PRIOR FILING DATE: 1995-03-10
PRIOR APPLICATION NUMBER: 09/782,161
PRIOR FILING DATE: 2001-02-14
PRIOR APPLICATION NUMBER: 08/790,776
PRIOR FILING DATE: 1997-01-30
PRIOR APPLICATION NUMBER: 60/010,803
PRIOR FILING DATE: 1996-01-30
PRIOR APPLICATION NUMBER: 08/799,180
PRIOR FILING DATE: 1997-02-12
PRIOR APPLICATION NUMBER: 60/011,618
PRIOR FILING DATE: 1996-02-13
PRIOR APPLICATION NUMBER: 08/803,609
PRIOR FILING DATE: 1997-02-21
PRIOR APPLICATION NUMBER: 60/011,985
PRIOR FILING DATE: 1996-02-21
PRIOR APPLICATION NUMBER: 08/971,050
PRIOR FILING DATE: 1997-11-14
PRIOR APPLICATION NUMBER: 60/031,038
PRIOR FILING DATE: 1996-11-18
PRIOR APPLICATION NUMBER: 08/975,985
PRIOR FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: 60/031,221
PRIOR FILING DATE: 1996-11-25
PRIOR APPLICATION NUMBER: 08/985,366
PRIOR FILING DATE: 1997-12-04
PRIOR APPLICATION NUMBER: 09/842,827
PRIOR FILING DATE: 2001-04-27
PRIOR APPLICATION NUMBER: 60/032,782
PRIOR FILING DATE: 1996-12-06
PRIOR APPLICATION NUMBER: 09/076,898
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/047,004
PRIOR FILING DATE: 1997-05-13
PRIOR APPLICATION NUMBER: 09/078,614
PRIOR FILING DATE: 1998-05-14
PRIOR APPLICATION NUMBER: 60/046,498
PRIOR FILING DATE: 1997-05-14
PRIOR APPLICATION NUMBER: 09/189,834
PRIOR FILING DATE: 1998-11-12
PRIOR APPLICATION NUMBER: 60/065,454

PRIOR FILING DATE: 1997-11-12
PRIOR APPLICATION NUMBER: 09/366,690
PRIOR FILING DATE: 1999-08-04
PRIOR APPLICATION NUMBER: 60/095,484
PRIOR FILING DATE: 1998-08-05
PRIOR APPLICATION NUMBER: 09/342,216
PRIOR FILING DATE: 1999-06-29
PRIOR APPLICATION NUMBER: 60/091,171
PRIOR FILING DATE: 1998-06-30
PRIOR APPLICATION NUMBER: 09/342,061
PRIOR FILING DATE: 1999-06-29
PRIOR APPLICATION NUMBER: 60/091,170
PRIOR FILING DATE: 1998-06-30
PRIOR APPLICATION NUMBER: 09/487,572
PRIOR FILING DATE: 2000-01-19
PRIOR APPLICATION NUMBER: 60/116,678
PRIOR FILING DATE: 1999-01-21
PRIOR APPLICATION NUMBER: 09/497,772
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: 60/118,784
PRIOR FILING DATE: 1999-02-05
PRIOR APPLICATION NUMBER: 09/504,577
PRIOR FILING DATE: 2000-02-15
PRIOR APPLICATION NUMBER: 60/120,434
PRIOR FILING DATE: 1999-02-16
PRIOR APPLICATION NUMBER: 09/517,011
PRIOR FILING DATE: 2000-03-01
PRIOR APPLICATION NUMBER: 60/122,447
PRIOR FILING DATE: 1999-03-02
PRIOR APPLICATION NUMBER: 09/628,857
PRIOR FILING DATE: 2000-07-27
PRIOR APPLICATION NUMBER: 60/145,956
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: 09/628,858
PRIOR FILING DATE: 2000-07-27
PRIOR APPLICATION NUMBER: 60/145,955
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: 09/741,033
PRIOR FILING DATE: 2000-12-21
PRIOR APPLICATION NUMBER: 09/773,518
PRIOR FILING DATE: 2001-02-02
PRIOR APPLICATION NUMBER: 60/171,621
PRIOR FILING DATE: 1999-12-23
PRIOR APPLICATION NUMBER: 09/796,765
PRIOR FILING DATE: 2001-03-02
PRIOR APPLICATION NUMBER: 09/867,682
PRIOR FILING DATE: 2001-05-31
PRIOR APPLICATION NUMBER: 60/187,016
PRIOR FILING DATE: 2000-03-06
PRIOR APPLICATION NUMBER: 60/225,269
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/225,629
PRIOR FILING DATE: 2000-08-15
NUMBER OF SEQ ID NOS: 55553
SEQ ID NO 55206
LENGTH: 273
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (42)..(42)
OTHER INFORMATION: n is equal to a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (56)..(56)
OTHER INFORMATION: n is equal to a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (69)..(69)
OTHER INFORMATION: n is equal to a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (125)..(126)
OTHER INFORMATION: n is equal to a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (176)..(176)

```
OTHER INFORMATION: n is equal to a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (199)..(199)
OTHER INFORMATION: n is equal to a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (244)..(244)
OTHER INFORMATION: n is equal to a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (247)..(247)
OTHER INFORMATION: n is equal to a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (260)..(260)
OTHER INFORMATION: n is equal to a,t,g, or c
US-09-912-292-55206

Query Match
Best Local Similarity 73.6%; Score 16.2; DB 7; Length 273;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 agcgcagctagctgcgtcct 22
    ||||| ||||| ||||| |||||
DB 236 AGCCCTACGGTAGCTGCTCT 216

RESULT 10
US-60-253-653-2481/C
; Sequence 2481, Application US/60253653
; GENERAL INFORMATION:
; APPLICANT: Glenn, Matthew
; TITLE OF INVENTION: Compositions isolated from ovine tissues
; FILE REFERENCE: 1057P
; CURRENT APPLICATION NUMBER: US/60/253,653
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 30124
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2481
; LENGTH: 277
; TYPE: DNA
; ORGANISM: Ovine
US-60-253-653-2481

Query Match
Best Local Similarity 73.6%; Score 16.2; DB 8; Length 277;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 agcgcagctagctgcgtcct 22
    ||||| ||||| ||||| |||||
DB 211 AGCCACGATGCTGCTCT 191

RESULT 11
US-60-252-833-32199/C
; Sequence 32199, Application US/60252833
; GENERAL INFORMATION:
; APPLICANT: Glenn, Matthew
; APPLICANT: Grigor, Murray R
; TITLE OF INVENTION: Compositions isolated from bovine
; FILE REFERENCE: 1052P2
; CURRENT APPLICATION NUMBER: US/60/252,833
; CURRENT FILING DATE: 2000-11-21
; NUMBER OF SEQ ID NOS: 43535
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 32199
; LENGTH: 326
; TYPE: DNA
; ORGANISM: Bovine
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(326)
```

```
OTHER INFORMATION: n = A,T,C or G
US-60-252-833-32199

Query Match
Best Local Similarity 73.6%; Score 16.2; DB 8; Length 326;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 agcgcagctagctgcgtcct 22
    ||||| ||||| ||||| |||||
DB 210 AGCCACGATGCTGCTCT 190

RESULT 12
US-60-253-652-13183/C
; Sequence 13183, Application US/60253652
; GENERAL INFORMATION:
; APPLICANT: Glenn, Matthew
; APPLICANT: Grigor, Murray R
; TITLE OF INVENTION: Compositions isolated from bovine
; FILE REFERENCE: 1055P2
; CURRENT APPLICATION NUMBER: US/60/253,652
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 29954
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13183
; LENGTH: 327
; TYPE: DNA
; ORGANISM: Bovine
US-60-253-652-13183

Query Match
Best Local Similarity 73.6%; Score 16.2; DB 8; Length 327;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 agcgcagctagctgcgtcct 22
    ||||| ||||| ||||| |||||
DB 215 AGCCACGATGCTGCTCT 195

RESULT 13
US-09-540-213-15508
; Sequence 15508, Application US/09540213
; GENERAL INFORMATION:
; APPLICANT: Selheimer, Jeffrey J.
; APPLICANT: Deleagane, Angelo M.
; APPLICANT: Stuart, Susan G.
; APPLICANT: Stuve, Laura L.
; APPLICANT: Mullahy, Sara J.
; APPLICANT: Naughton, Rebecca E.
; TITLE OF INVENTION: POLYNUCLEOTIDES OF LIVER AND PANCREAS
; FILE REFERENCE: PD-1031 CIP
; CURRENT APPLICATION NUMBER: US/09/540,213
; CURRENT FILING DATE: 2000-03-31
; PRIOR APPLICATION data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 53844
; SEQ ID NO 15508
; LENGTH: 332
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: hu01083786
US-09-540-213-15508

Query Match
Best Local Similarity 73.6%; Score 16.2; DB 7; Length 332;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 cagcgcagctagctgcgtcct 21
```

Db 85 cagccgagcagctgcctcc 105

```

RESULT 14
US-60-253-652-14207/c
: Sequence 14207, Application US/60253652
: GENERAL INFORMATION:
: APPLICANT: Glenn, Matthew
: TITLE OF INVENTION: Compositions isolated from bovine
: FILE REFERENCE: 1055P2
: CURRENT APPLICATION NUMBER: US/60/253,652
: NUMBER OF SEQ ID NOS: 29954
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 14207
: LENGTH: 345
: TYPE: DNA
: ORGANISM: Bovine
US-60-253-652-14207
    
```

Query Match 73.6%; Score 16.2; DB 8; Length 345;
 Best Local Similarity 85.7%; Pred. No. 1.7e+02;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 agccgagctagctgcctcct 22
 ||||| ||| |||||
 Db 191 AGCCGAGCTAGCTGCCTCCT 171

```

RESULT 15
US-60-252-833-19140/c
: Sequence 19140, Application US/60252833
: GENERAL INFORMATION:
: APPLICANT: Glenn, Matthew
: TITLE OF INVENTION: Compositions isolated from bovine
: FILE REFERENCE: 1052P2
: CURRENT APPLICATION NUMBER: US/60/252,833
: CURRENT FILING DATE: 2000-11-21
: NUMBER OF SEQ ID NOS: 43535
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 19140
: LENGTH: 369
: TYPE: DNA
: ORGANISM: Bovine
US-60-252-833-19140
    
```

Query Match 73.6%; Score 16.2; DB 8; Length 369;
 Best Local Similarity 85.7%; Pred. No. 1.8e+02;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 agccgagctagctgcctcct 22
 ||||| ||| |||||
 Db 208 AGCCGAGCTAGCTGCCTCCT 188

Search completed: September 21, 2001, 02:39:25
 Job time: 15051 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 21, 2001, 02:42:32 : Search time 330.77 Seconds
(without alignments)
34.169 Million cell updates/sec

Title: US-09-138-735-8

Perfect score: 18

Sequence: 1 tgcagcagcgcgagcaagt 18

Scoring table: IDENTITY_NUC

Searched: Gapop 10.0 , Gapext 1.0

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

N_Geneseq_0601 :
1: /cgnl_9/gcgdata/geneseq/geneseqn/NA1980.DAT : *
2: /cgnl_9/gcgdata/geneseq/geneseqn/NA1981.DAT : *
3: /cgnl_9/gcgdata/geneseq/geneseqn/NA1982.DAT : *
4: /cgnl_9/gcgdata/geneseq/geneseqn/NA1983.DAT : *
5: /cgnl_9/gcgdata/geneseq/geneseqn/NA1984.DAT : *
6: /cgnl_9/gcgdata/geneseq/geneseqn/NA1985.DAT : *
7: /cgnl_9/gcgdata/geneseq/geneseqn/NA1986.DAT : *
8: /cgnl_9/gcgdata/geneseq/geneseqn/NA1987.DAT : *
9: /cgnl_9/gcgdata/geneseq/geneseqn/NA1988.DAT : *
10: /cgnl_9/gcgdata/geneseq/geneseqn/NA1989.DAT : *
11: /cgnl_9/gcgdata/geneseq/geneseqn/NA1990.DAT : *
12: /cgnl_9/gcgdata/geneseq/geneseqn/NA1991.DAT : *
13: /cgnl_9/gcgdata/geneseq/geneseqn/NA1992.DAT : *
14: /cgnl_9/gcgdata/geneseq/geneseqn/NA1993.DAT : *
15: /cgnl_9/gcgdata/geneseq/geneseqn/NA1994.DAT : *
16: /cgnl_9/gcgdata/geneseq/geneseqn/NA1995.DAT : *
17: /cgnl_9/gcgdata/geneseq/geneseqn/NA1996.DAT : *
18: /cgnl_9/gcgdata/geneseq/geneseqn/NA1997.DAT : *
19: /cgnl_9/gcgdata/geneseq/geneseqn/NA1998.DAT : *
20: /cgnl_9/gcgdata/geneseq/geneseqn/NA1999.DAT : *
21: /cgnl_9/gcgdata/geneseq/geneseqn/NA2000.DAT : *
22: /cgnl_9/gcgdata/geneseq/geneseqn/NA2001.DAT : *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18	100.0	18	17	AAAT7312
2	18	100.0	18	20	AAAB4094
3	18	100.0	18	20	AAAT7310
4	18	100.0	3402	20	AAAB4092
5	18	100.0	11878	19	AAV30199
6	18	100.0	11883	19	AAV30198
7	16.4	91.1	876	8	AAAT7094
8	16.4	91.1	36519	19	AAV22141
9	15.4	85.6	20	21	AAAI4833
10	15.4	85.6	39	20	AAAX0696
11	15.4	85.6	240	21	AAAS3346

12	15.4	85.6	303	16	AAAT18914
13	15.4	85.6	328	21	AAAC12718
14	15.4	85.6	341	21	AAAC03564
15	15.4	85.6	422	21	AAAC79295
16	15.4	85.6	448	21	AAAB95498
17	15.4	85.6	457	21	AAAB9684
18	15.4	85.6	478	21	AAAB9684
19	15.4	85.6	502	21	AAAB9683
20	15.4	85.6	1269	21	AAAC42299
21	15.4	85.6	1276	21	AAAI5086
22	15.4	85.6	1381	12	AAAI4472
23	15.4	85.6	1558	21	AAAC49642
24	15.4	85.6	1837	21	AAAC77971
25	15.4	85.6	1955	21	AAZ99588
26	15.4	85.6	2000	16	AAAB7925
27	15.4	85.6	2030	21	AAAC98843
28	15.4	85.6	2177	21	AAZ57853
29	15.4	85.6	2204	21	AAAC77452
30	15.4	85.6	2214	19	AAV22682
31	15.4	85.6	3000	21	AAAB9899
32	15.4	85.6	3183	20	AAAZ6611
33	15.4	85.6	3212	22	AAAC84373
34	15.4	85.6	3331	19	AAV22683
35	15.4	85.6	3331	20	AAZ32021
36	15.4	85.6	3331	22	AAAC90078
37	15.4	85.6	3426	20	AAZ23819
38	15.4	85.6	5197	16	AAQ79038
39	15.4	85.6	5228	20	AAV81384
40	15.4	85.6	6115	21	AAZ46519
41	15.4	85.6	34094	20	AAZ30163
42	15.4	85.6	273254	21	AAAC81914
43	15.4	83.3	362	22	AAAF93718
44	15.4	83.3	426	20	AAAB08739
45	15.4	83.3	1065	21	AAAF14537

ALIGNMENTS

RESULT 1
ID AAT7312
AAAT7312 standard; cDNA: 18 BP.
AC AAT7312;
DT 26-NOV-1996 (first entry)
DE T. cruzi epimastigotic PTC100t antigen primer corresp. to bases 1442-59.
KW Antigen; Trypanosoma cruzi; epimastigote; serum; Chagas disease; probe;
KW Primer; PCR; polymerase chain reaction; amplification; antibody; ss.
OS Synthetic.
PM FR2723589-A1.
PD 16-FEB-1996.
PF 12-AUG-1994; 94FR-0010132.
PR 12-AUG-1994; 94FR-0010132.
XX (INMR) BTO MERITUX.
PA Jollivet M, Lesenechal M, Paranhos-Bacala G;
PI WPI; 1996-190287/20.
DR New nucleic acid encoding Trypanosoma cruzi epimastigotic antigen -
PT useful for diagnosis, monitoring and therapy of Chagas disease
PS Claim 29; Page 36; 55pp; French.

CC The primers AAT7311-5 were used to PCR amplify the sequence encoding a
CC novel isolated antigenic protein from Trypanosoma cruzi epimastigotes,
CC designated PTC100t (AAT7310). The primers, derived from the sequences
CC of a 594 and 1041 bp fragment of PTC100t, amplified the gene as 3
CC fragments. This primer corresponds to nucleotides 1442-59 of the pTC100t
CC sequence, derived from the 594 bp fragment. The 594 bp fragment was
CC isolated from a T. cruzi genomic expression library in lambda gtl1, using
CC a mixture of sera from patients with Chagas disease. It corresponds to
CC nucleotides 1232-1825 of PTC100t. The 1041 bp fragment was isolated from
CC a lambda gtl10 library using the 594 bp fragment as a probe. The protein,
CC or antibodies raised against it, can be used in the detection and
CC monitoring of T. cruzi infection i.e. Chagas disease.

XX Sequence 18 BP; 5 A; 4 C; 7 G; 2 T; 0 other;

Query Match 100.0%; Score 18; DB 17; Length 18;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 tgcagcagcgcgcagaagt 18
|||||
DB 1 tgcagcagcgcgcagaagt 18

RESULT 2

AAx84094
ID AAx84094 standard; DNA; 18 BP.

AC AAx84094;

DT 27-AUG-1999 (first entry)

DE PCR primer for T. cruzi PTC40 coding sequence.

KW PTC40; Tc40; infection; diagnosis; immune complex; antigenic determinant;
KW therapy; antibody; PCR primer; ss.

OS Synthetic.

OS Trypanosoma cruzi.

PN WO9929867-A1.

PD 17-JUN-1999.

PF 10-DEC-1998; 98WO-IB01987.

PR 10-DEC-1997; 97US-0988242.

PA (INMR) BIO MERIEUX.

PI Jolivet M, Lesenechal M, Mandrand B, Paranhos-Baccala G;

DR WPI: 1999-394978/73.

PT New Trypanosoma cruzi antigen

PS Disclosure; Page 21; 65pp; English.

XX This sequence is a PCR primer for DNA encoding the Trypanosoma cruzi
CC PTC40 protein of the invention. The PTC40 antigenic determinant is
CC useful as a reagent for detection and/or monitoring of Trypanosoma cruzi
CC infection from samples including blood serum or plasma, urine, saliva, or
CC tears, by contacting with the sample and detecting an immune complex. The
CC PTC40 antigenic determinant, the vector, expression cassette, cell or
CC antibody are useful for treatment or prevention (vaccine) of a
CC Trypanosoma cruzi infection in a man or animal. Current Trypanosoma cruzi
CC antigens are obtained from protein fractions of the noninfectious stage
CC of the parasite, and these do not allow sufficient production of antigens
CC for use in reliable serological diagnostic tests. The strain to strain
XX polymorphism reduces reliability of the tests.

XX Sequence 18 BP; 5 A; 4 C; 7 G; 2 T; 0 other;

Query Match 100.0%; Score 18; DB 20; Length 18;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 tgcagcagcgcgcagaagt 18
|||||
DB 1 tgcagcagcgcgcagaagt 18

RESULT 3

AAT7310/C
ID AAT7310 standard; cDNA; 3402 BP.

AC AAT7310;

DT 26-NOV-1996 (first entry)

DE Trypanosoma cruzi epimastigotic PTC100t antigen gene.

KW Antigen; Trypanosoma cruzi; epimastigote; serum; Chagas disease; probe;
KW primer; PCR; polymerase chain reaction; amplification; antibody; ds.

OS Trypanosoma cruzi.

FT Key Location/Qualifiers

FT CDS 266..3013

FT /product= PTC100t epimastigotic antigen

PN FR2723589-A1.

PD 16-FEB-1996.

PF 12-AUG-1994; 94FR-0010132.

PR 12-AUG-1994; 94FR-0010132.

PA (INMR) BIO MERIEUX.

PI Jolivet M, Lesenechal M, Paranhos-Baccala G;

DR WPI: 1996-190287/20.

PT P-PSDB; AAR91615.

PT New nucleic acid encoding Trypanosoma cruzi epimastigotic antigen -
PT useful for diagnosis, monitoring and therapy of Chagas disease

PS Claim 1; Page 24-26; 55pp; French.

XX This is the nucleotide sequence encoding a novel isolated antigenic
CC protein from Trypanosoma cruzi epimastigotes, designated PTC100t.
CC The clone Tc50 was isolated from a T. cruzi genomic expression library in
CC lambda gtl1, using a mixture of sera from patients with Chagas disease;
CC clone Tc50 contained an 594 bp insert corresp. to nucleotides 1232-1825
CC of this sequence. The Tc50 sequence was subsequently used to probe a
CC Southern blot of restriction enzyme digested T. cruzi DNA and also screen
CC a lambda gtl10 library to isolate a 1041 bp EcoRI fragment corresp. to
CC nucleotides 1403-2443 of PTC100t. Primers (AAT7311-5) were synthesised
CC based on the sequences of the 594 and 1041 bp fragments and used to
CC amplify the PTC100t clone as 3 fragments from cDNA derived from mRNA
CC purified from T. cruzi epimastigotes. The protein or antibodies raised
CC against it can be used in the detection and monitoring of T. cruzi
XX infection i.e. Chagas disease.

XX Sequence 3402 BP; 889 A; 818 C; 958 G; 737 T; 0 other;

Query Match 100.0%; Score 18; DB 17; Length 3402;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 tgcagcagcgagcaagt 18
 |||||||
 DB 1459 TGCAGCAGCGGACAGACT 1442

RESULT 4
 AAX84092/c
 ID AAX84092 standard; cDNA; 3402 BP.
 XX
 AC AAX84092;
 XX
 DT 27-AUG-1999 (first entry)
 XX
 DE T. cruzi PTC40 coding sequence.
 XX
 KM PTC40; Tc40; infection; diagnosis; immune complex; antigenic determinant;
 KW therapy; antibody; ds.
 XX
 OS Trypanosoma cruzi.
 XX
 PN WO9929867-A1.
 PD 17-JUN-1999.
 XX
 PF 10-DEC-1998; 98WO-IB01987.
 XX
 PR 10-DEC-1997; 97US-0988242.
 XX
 PA (INMR) BIO MERIEUX.
 XX
 PI Jollivet M, Lesenechal M, Mandrand B, Paranhos-Baccala G;
 DR WPI: 1999-394978/33.
 DR P-PSDB; AAV22124.
 XX
 PT New Trypanosoma cruzi antigen
 XX
 PS Claim 1; Page 52-56; 65pp; English.

XX This sequence encodes the Trypanosoma cruzi PTC40 protein of the
 CC invention, and is designated Tc40. The PTC40 antigenic determinant is
 CC useful as a reagent for detection and/or monitoring of Trypanosoma cruzi
 CC infection from samples including blood serum or plasma, urine, saliva, or
 CC tears, by contacting with the sample and detecting an immune complex. The
 CC PTC40 antigenic determinant, the vector, expression cassette, cell or
 CC antibody are useful for treatment or prevention (vaccine) of a
 CC Trypanosoma cruzi infection in a man or animal. Current Trypanosoma cruzi
 CC antigens are obtained from protein fractions of the noninfectious stage
 CC of the parasite, and these do not allow sufficient production of antigens
 CC for use in reliable serological diagnostic tests. The strain to strain
 CC polymorphism reduces reliability of the tests.
 XX
 SQ Sequence 3402 BP; 888 A; 821 C; 956 G; 737 T; 0 other;

Query Match 100.0%; Score 18; DB 20; Length 3402;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 tgcagcagcgagcaagt 18
 |||||||
 DB 1459 TGCAGCAGCGGACAGACT 1442

RESULT 5
 AAV30199
 ID AAV30199 standard; DNA; 11878 BP.
 XX
 AC AAV30199;
 XX
 DT 14-SEP-1998 (first entry)
 XX
 DE Protein kinase catalytic subunit gene.

XX
 KW Severe combined immunodeficiency disease; SCID; horse; diagnosis;
 KW DNA-dependent protein kinase; ds.
 XX
 OS Equus caballus.
 XX
 PN WO9821367-A1.
 XX
 PD 22-MAY-1998.
 XX
 PF 14-NOV-1997; 97WO-US21066.
 XX
 PR 15-NOV-1996; 96US-0031261.
 XX
 PA (TEXA) UNIV TEXAS SYSTEM.
 XX
 PI Meeks K;
 DR WPI: 1998-297967/26.
 XX
 PT DNA-dependent protein kinase catalytic subunit - useful for
 PT determining equine severe combined immunodeficiency alleles
 XX
 PS Disclosure; Page 70-78; 98pp; English.

XX This isolated DNA molecule encodes an equine DNA-dependent protein
 CC kinase catalytic subunit (DNA-PKcs). A claimed method of
 CC identifying an Arabian horse that is a carrier of equine severe
 CC combined immunodeficiency (SCID) comprises determining whether the
 CC horse has a mutation in a SCID determinant region of the DNA-PKcs
 CC gene (see also AAV30196 and AAV30197). Sequence analysis of DNA-PKcs
 CC genes from normal and SCID equine fibroblasts shows that a 5 bp
 CC deletion is present in SCID foals at a site that corresponds to
 CC nucleotide 9454 of the 12,381 nucleotide coding sequence of the
 CC human transcript. This results in premature termination of the
 CC DNA-PKcs at amino acid 3160 (see AAW56642). Oligonucleotide probes
 CC (see AAV30194 and AAV30195) are provided that precisely span the SCID
 CC determinant region of the DNA-PKcs gene, and which are diagnostic
 CC for the normal and SCID alleles, respectively. Methods are also
 CC provided for identifying for differentiating SCID homozygotes,
 CC heterozygotes and normal horses.
 XX
 SQ Sequence 11878 BP; 3586 A; 2323 C; 2668 G; 3301 T; 0 other;

Query Match 100.0%; Score 18; DB 19; Length 11878;
 Best Local Similarity 100.0%; Pred. No. 45;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 tgcagcagcgagcaagt 18
 |||||||
 DB 6429 tgcagcagcgagcaagt 6446

RESULT 6
 AAV30198
 ID AAV30198 standard; DNA; 11883 BP.
 XX
 AC AAV30198;
 XX
 DT 14-SEP-1998 (first entry)
 XX
 DE Protein kinase catalytic subunit gene.
 XX
 KW Severe combined immunodeficiency disease; SCID; horse; diagnosis;
 KW DNA-dependent protein kinase; ds.
 XX
 OS Equus caballus.
 XX
 PN WO9821367-A1.
 XX
 PD 22-MAY-1998.
 XX

PF 14-NOV-1997; 97WO-US21066.
 XX 15-NOV-1996; 96US-0031261.
 XX (TEXA) UNIV TEXAS SYSTEM.
 XX
 PI Weeks K;
 DR WPI: 1998-297967/26.
 DR P-PSDB; AAW56642.
 XX
 PT DNA-dependent protein kinase catalytic subunit - useful for
 PT determining equine severe combined immunodeficiency alleles
 XX
 XX Claim 1; Page 39-44; 98pp; English.
 PS
 CC This isolated DNA molecule encodes a DNA-dependent protein kinase
 CC catalytic subunit (DNA-PKcs, see AAW56642) found in Arabian horses.
 CC A claimed method of identifying an Arabian horse that is a carrier
 CC of equine severe combined immunodeficiency (SCID) comprises
 CC determining whether the horse has a mutation in a SCID determinant
 CC region of the DNA-PKcs gene (see also AAV30196 and AAV30197). Sequence
 CC analysis of DNA-PKcs genes from normal and SCID equine fibroblasts
 CC shows that a 5 bp deletion is present in SCID foals at a site that
 CC corresponds to nucleotide 9434 of the 12,381 nucleotide coding
 CC sequence of the human transcript. This results in premature
 CC termination of the DNA-PKsc at amino acid 3160. Oligonucleotide
 CC probes (see AAV30194 and AAV30195) are provided that precisely span the
 CC SCID determinant region of the DNA-PKcs gene, and which are
 CC diagnostic for the normal and SCID alleles, respectively. Methods
 CC are also provided for identifying for differentiating SCID
 CC homozygotes, heterozygotes and normal horses. A claimed plasmid
 CC contains the DNA-PKcs DNA and regulatory elements necessary for
 CC expression of the DNA in a recombinant cell.
 SQ Sequence 11883 BP; 3588 A; 2324 C; 2669 G; 3302 T; 0 other;

Query Match 100.0%; Score 18; DB 19; Length 11883;
 Best Local Similarity 100.0%; Pred. No. 45;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgcagcagcgagcaagt 18
 ||||||||||||||||
 DB 6432 tgcagcagcgagcaagt 6449

RESULT 7
 AAN70094
 ID AAN70094 standard; DNA; 876 BP.
 XX
 AC AAN70094;
 XX
 DT 24-APR-1991 (first entry)
 XX
 DE Sequence of region in front of the neutral protease gene in the
 DE Bacillus subtilis strain BGSC 1A1.
 XX
 DE Promoter; Bacillus subtilis; pSM127; ss.
 KW
 KW Bacillus subtilis.
 XX
 OS
 XX
 XX Key Location/Qualifiers
 FH 452..517 /tag= a
 FT misc_diff /tag= a
 FT note= "absent from nprR2"
 FT repeat_region 510..577 /tag= b
 FT /rpt_type= INVERTED
 FT /note= "3 pairs of inverted repeats"
 FT promoter 571..599 /tag= c
 FT /label= P1

FT promoter 574..598
 FT /tag= d
 FT /label= P2
 XX
 PN EP213085-A.
 XX
 PD 04-MAR-1987.
 XX
 PF 30-JUN-1986; 86EP-0830185.
 XX
 PR 10-JUL-1985; 85IT-0021507.
 XX
 PA (ENTE) ENRICECERHE SPA.
 XX
 PI Toma S, Del Bue M, Grandi G, Mele A;
 XX
 DR WPI: 1987-058527/09.
 XX
 PT Recombinant DNA cloning vector - used for the expression and
 PT secretion of heterologous gene prods. in Bacillus subtilis
 PS
 CC Disclosure; Fig 2B; 17pp; English.
 XX
 CC The nprR2 region is differentiated from that of B. subtilis BGSC 1A1
 CC in that there is a deletion of 66 base pairs at a distance of 124
 CC bp from the origin of the gene, (see AAN70093, N70094). A new cloning
 CC vector based on pSM127 is claimed in which the heterologous DNA is under
 CC the control of the promoter regions, the centre of recognition of
 CC the ribosomes, nprR2 and the signal sequence of the neutral protease.
 XX
 SQ Sequence 876 BP; 289 A; 185 C; 159 G; 243 T; 0 other;

Query Match 91.1%; Score 16.4; DB 8; Length 876;
 Best Local Similarity 94.4%; Pred. No. 1.9e+02;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 tgcagcagcgagcaagt 18
 ||||||||||||||||
 DB 63 tgcagcagcgagcaagt 80

RESULT 8
 AAV22141
 ID AAV22141 standard; cDNA; 36519 BP.
 XX
 AC AAV22141;
 XX
 DT 17-AUG-1998 (first entry)
 XX
 DE Chimpanzee adenovirus C68 genomic sequence.
 XX
 DE genomic sequence; viral genome; virus; gene therapy; treatment;
 KW acquired defects; inherited defects; genetic engineering; vector;
 KW in vitro production; recombinant protein; ds.
 XX
 OS Chimpanzee adenovirus C68.
 XX
 FH Key Location/Qualifiers
 FT CDS 480..1521 /tag= a
 FT /note= "E1A gene"
 FT CDS 1560..3956 /tag= b
 FT /note= "E1B gene"
 FT CDS complement(23370..21787) /tag= c
 FT /note= "E2A gene"
 FT CDS complement(10346..3957) /tag= d
 FT /note= "E2B gene"
 FT CDS 26806..31877 /tag= e


```

FT      CDS                               /note= "E3 gene"
FT      CDS                               /complement(36193..33486)
FT      CDS                               /*tag= f
FT      CDS                               /note= "E4 gene"
FT      CDS                               10823..13817
FT      CDS                               /*tag= g
FT      CDS                               /note= "L1 gene"
FT      CDS                               13884..17431
FT      CDS                               /*tag= h
FT      CDS                               /note= "L2 gene"
FT      CDS                               17480..21804
FT      CDS                               /*tag= i
FT      CDS                               /note= "L3 gene"
FT      CDS                               23389..27439
FT      CDS                               /*tag= j
FT      CDS                               /note= "L4 gene"
FT      CDS                               32134..33502
FT      CDS                               /*tag= k
FT      CDS                               /note= "L5 gene"

XX      PN      W09810087-A1.
XX      PD      12-MAR-1998.
XX      PE      04-SEP-1997; 97MO-US15694.
XX      PR      06-SEP-1996; 96US-0024700.
XX      PA      (UYPE-) UNIV PENNSYLVANIA.
XX      PI      Farina SF, Fisher KJ, Wilson JM;
XX      DR      WPI; 1998-193635/17.
XX      PT      Chimpanzee adenovirus vector - useful in gene therapy and genetic
XX      PS      engineering in general
XX      PS      Disclosure; Pages 70-89; 116pp; English.
XX      CC      The sequence is that of the chimpanzee adenovirus C68. It may be
XX      CC      used in the production of a vector comprising it and a selected
XX      CC      heterologous gene operatively linked to regulatory sequences
XX      CC      directing its expression in a heterologous cell.
XX      CC      Such recombinant adenoviruses are useful as vectors in gene
XX      CC      therapy and genetic engineering in general, especially for
XX      CC      the treatment of acquired or genetically inherited defects.
XX      CC      The viruses are also useful for the in vitro production of
XX      CC      recombinant proteins of interest.
XX      SQ      Sequence 36519 BP; 7809 A; 10790 C; 10623 G; 7297 T; 0 other;

Query Match          91.1%; Score 16.4; DB 19; Length 36519;
Best Local Similarity 94.4%; Pred. No. 2.2e+02;
Matches 17; Conservatively 0; Mismatches 1; Indels 0; Gaps 0;

QY      1      tgcagcagcgcgcaagaat 18
          ||||||||| |||||||
Db      31662 tgcagcagcgcgcaagaat 31679

RESULT      9
AAAI4833/c
ID      AAI4833 standard; DNA; 20 BP.
XX
XX      AAI4833;
XX
XX      08-AUG-2000 (first entry)
XX
XX      PCR primer HG52.165F for human G-protein coupled receptor HG52 cDNA.
XX
XX      Human; G-protein coupled receptor; HG52; Immune system; thrombin;
XX      fibrinogen; fibrin; clotting factor; procoagulant; platelet activation;

```

KW		chemotaxis; mitogenesis; PCR primer; ss.
XX		
OS	Homo sapiens.	
XX		
PN	WC200020438-A1.	
XX		
PD	13-APR-2000.	
XX		
PF	29-SEP-1999;	99WO-US22634.
XX		
PR	02-OCT-1998;	98US-0102958.
XX		
PA	(MERI) MERCK & CO INC.	
XX		
P1	Liu Q, McDonald TP, Wang R;	
DR	WPI; 2000-317696/27.	
XX		
PT	New recombinant DNA encoding a G-protein coupled receptor designated	
PT	HG52 is useful to find modulators of thrombin effects and shows	
PT	homology to thrombin receptors -	
XX		
PS	Example 1; Page 23; 39pp; English.	
XX		
CC	PCR primers AAA14831-44 were used to amplify cDNA encoding a human	
CC	G-protein coupled receptor designated HG52. HG52 is a member of	
CC	the rhodopsin family. The HG52 RNA is widely expressed in humans as	
CC	a transcript of about 4.5. kb, especially in cells of the immune	
CC	system. The HG52 DNA can be used in chromosomal mapping studies, and	
CC*	to identify individuals carrying a disease-carrying gene. Agonists	
CC	and antagonists of HG52 will be useful as modulators of the effects	
CC	of thrombin, including conversion of fibrinogen to fibrin in plasma,	
CC	activation of clotting factors V, VIII, XIII and protein C, the	
CC	procoagulant function of platelets and endothelial cells, stimulation	
CC	of platelet activation, chemotaxis of monocytes and lymphocytes, and	
CC	mitogenesis of lymphocytes and mesenchymal cells such as vascular	
CC	smooth muscle cells, fibroblasts and epithelial cells.	
XX		
SQ	Sequence 20 BP; 1 A; 8 C; 4 G; 7 T; 0 other;	
OY	2 gcacgacgcgcagaagt 18	
DB	20 GAAGCAGCGCGCAGAAGT 4	
RESULT 10		
ID	AAAX06966	
ID	AAAX06966 standard; DNA; 39 BP.	
XX		
AC	AAAX06966;	
XX		
DT	10-MAY-1999 (first entry)	
XX		
DE	Bacillus agaradhaerens primer xyloglucanase.upper.SacII.	
XX		
KW	Xyloglucanase; detergent; PCR; primer; ss.	
XX		
OS	Synthetic.	
OS	Bacillus agaradhaerens.	
XX		
PN	W09902663-A1.	
XX		
PD	21-JAN-1999.	
XX		
PF	01-JUL-1998;	98WO-DK00290.
XX		
XX	24-OCT-1997;	97DK-0001213.
PR	07-JUL-1997;	97DK-0000822.
PR		

XX (NOVO) NOVO-NORDISK AS.
 XX
 CC Bjornvad ME, Jorgensen PL, Outtrup H, Schuelein M;
 XX
 XX WPI; 1999-120866/10.
 DR
 XX
 PT New enzyme preparation comprising a xyloglucanase with an activity
 of 50 % at pH 7 - useful for improving the properties of cellulosic
 fibres, yarn, (non)woven fabric, and rating hemp, jute, flax and
 PT linen fibres
 XX
 PS Example 6; Page 63; 87pp; English.
 CC
 CC Primers xyloglucanase: upper: SalII and xyloglucosidase: lower: NotI
 CC (see AA06967) were used for the PCR amplification of Bacillus
 CC agaradhaerens NCIMB 40482 chromosomal DNA. The primers incorporate
 CC SalII and NotI sites into the amplified DNA. The PCR product was
 CC ligated into plasmid pMO1944, a pUB110 derivative containing
 CC elements making it propagatable in Bacillus subtilis. The B.
 CC agaradhaerens xyloglucanase (see AA08463) was expressed in
 CC transformed B. subtilis PL2954 cells. The enzyme is useful for
 CC improving the properties of cellulosic fabrics and in detergent
 CC compositions.
 CC
 XX
 SQ Sequence 39 BP; 9 A; 12 C; 10 G; 8 T; 0 other;

Query Match 85.6%; Score 15.4; DB 20; Length 39;
 Best Local Similarity 94.1%; Pred. No. 4.4e+02;
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 tgcagcagcgccagcaag 17
 ||||| |||||
 Db 6 tgcagcagcgccagcaag 22

RESULT 11

AAA35346
 ID AAA35346 standard; DNA; 240 BP.

AC AAA35346;
 XX

DT 25-JUL-2000 (first entry)

DE Myrtaceae microsatellite sequence scu0737r including flanking regions.

KM Myrtaceae: microsatellite: isolation; genotyping; plant; tea tree;
 KM breeding; Melaleuca alternifolia: broad-spectrum germicidal oil;
 KM pharmaceutical; cosmetic; identification; detection; ds.

XX
 XX Myrtaceae sp.

OS
 XX

PN WO200017341-A1.

PD 30-MAR-2000.

PF 23-SEP-1999; 99WO-AU00820.

PR 23-SEP-1998; 98AU-0006099.

PR 16-FEB-1999; 99AU-0008718.

PA (BUST-) BUSINESS & RES MANAGEMENT PTY LTD.

PI Rossetto M, McLauchlan A, Harriss FCL, Henry RJ, Baverstock PR;
 PI Lee LS, Maguire TL, Edwards KJ;

XX WPI; 2000-292840/25.

XX Isolating microsatellites from Myrtaceae, useful for genotyping,
 PT particularly in breeding programs for tea tree, by reacting plant
 PT nucleic acid with immobilized oligonucleotides
 XX

PS Claim 8; Fig 35; 100pp; English.

XX A method has been developed of isolating a microsatellite (MS) from
 CC nucleic acid extract of a plant of Myrtaceae family. The method
 CC comprises: (i) treating the extract with one or more immobilised,
 CC single-stranded oligonucleotides (ON) having a consensus MS repeat
 CC sequence (MSRS) or its complement; (ii) washing under specified
 CC stringency conditions; (iii) eluting nucleic acid bound to ON; and
 CC (iv) sequencing the eluted nucleic acids to identify those containing
 CC an MSRS. Microsatellites (MS) isolated by the method, specifically
 CC from Melaleuca alternifolia (the tea tree, a source of a broad-spectrum
 CC germicidal oil, useful in pharmaceuticals and cosmetics), are useful as
 CC genotyping markers, particularly for breeding plants that produce the
 CC oil in higher yield or of better quality. Primers based on MS are
 CC useful for both inter- and intra-species genotyping. The selected
 CC washing conditions improve efficiency of recovery of microsatellites
 CC (MS) and reduce the number of washing stages required. Particularly
 CC about 86% of recovered sequence contain an MS repeat sequence.
 CC compared with 50-70% when the conventional washing procedure is
 CC followed. AAA35313 to AAA35357, and AAA35362 to AAA35375 represent
 CC nucleotide sequences from the present invention which contain
 CC microsatellite sequences. AAA35358 to AAA35361 represent oligonucleotide
 CC PCR primers used for identifying Myrtaceae microsatellite sequences.
 CC
 XX
 SQ Sequence 240 BP; 52 A; 63 C; 67 G; 54 T; 4 other;

Query Match 85.6%; Score 15.4; DB 21; Length 240;
 Best Local Similarity 94.1%; Pred. No. 4.8e+02;
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 tgcagcagcgccagcaag 17
 ||||| |||||
 Db 107 tgcagcagcgccagcaag 123

RESULT 12

AAT18914
 ID AAT18914 standard; DNA; 303 BP.

AC AAT18914;
 XX

DT 17-JAN-1997 (first entry)

DE DNA encoding spider dragline variant, DP-1B.16 monomer.

KM Spider: dragline protein; variant; monomer; polymer;
 KM fibre forming region; Spideroin 1; Nephila clavipes; Dpl; mimic;
 KM DP-1A analogue; fibre; high tensile strength; elasticity; clothing;
 KM rope; surgical suture; implant; reinforcement; film; coating; ss.

XX
 XX Synthetic.

OS
 XX

PN WO9429450-A2.

PD 22-DEC-1994.

PF 15-JUN-1994; 94WO-US06689.

PR 15-JUN-1993; 93US-0077600.

PA (DUPO) DU PONT DE NEMOURS & CO E I.

PI Fahnestock SR;

DR WPI; 1995-036479/05.

DR P-PSDB; AAR99056.

XX New synthetic variants of spider dragline protein - for making
 PT fibres useful as clothing, surgical silk, plastic reinforcement
 PT etc., also related DNA, vectors and transformed cells
 XX
 PS Claim 8; Page 122; 168pp; English.

XX This sequence encodes a synthetic spider dragline variant monomer,
 CC DP-1B.16. The sequence of the DP-1B.16 polymer is given in AAB99057.
 CC The polypeptide monomer is a variant based on a consensus sequence
 CC derived from the fibre forming regions of spider dragline protein,
 CC esp. the natural protein 1 (Spidroin 1) from Nephila clavipes. This
 CC DNA sequence may be used in the recombinant production of the
 CC variant protein in a recombinant host, e.g. *E. coli* or *Bacillus*
 CC subtilis. Synthetic analogues of DPI were designed to mimic the
 CC repeating consensus sequence of the natural protein and the pattern
 CC of variation among individual repeats. DP-1B analogues were designed
 CC by reordering the four repeats within the monomer of DPIA. This
 CC monomer exhibits all of the regularities of (1)-(5) below. In addition,
 CC it exhibits a regularity of the natural sequence which is not shared by
 CC DP-1A, namely that a repeat in which both GYG and GRG are deleted is
 CC generally preceded by a repeat lacking the entire poly-alanine repeat,
 CC with one intervening repeat. The sequence of DP-1B matches the natural
 CC sequence more closely over a more extended segment than does DP-1A. The
 CC individual repeats differ from the consensus according to the pattern:
 CC (1) the poly-alanine sequence varies in length from 0-7
 CC residues; (2) when the entire poly-alanine sequence is deleted,
 CC so also is the surrounding sequence encompassing AGRGSLGGQAGANGC;
 CC (3) aside from the poly-alanine sequence, deletions usually
 CC encompass integral multiples of three consecutive residues;
 CC (4) deletion of GYG is generally accompanied by deletion of GRG
 CC in the same sequence; and
 CC (5) a repeat in which the entire poly-alanine sequence is
 CC deleted is generally preceded by a repeat containing six alanine
 CC residues.
 CC The proteins may be used to produce fibres of high tensile strength and
 CC elasticity, suitable for clothing, rope, surgical sutures, biomaterials
 CC for implants, plastic reinforcements, films, coatings, etc.
 CC
 XX Sequence 303 BP; 34 A; 67 C; 140 G; 62 T; 0 other;

Query Match 85.6%; Score 15.4; DB 16; Length 303;
 Best Local Similarity 94.1%; Pred. No. 4.8e+02;
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 tgcagcagcgagcagaag 17
 |||||
 Db 249 tgcagcagcgagcagaag 265

RESULT 13
 AAC12718/c

ID AAC12718 standard; cDNA; 328 BP.

AC AAC12718;

DT 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 16793.

KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping; ss.

OS Homo sapiens.

PN EP1033401-A2.

PD 06-SEP-2000.

PF 21-FEB-2000; 2000EP-0200610.

PR 26-FEB-1999; 99US-0122487.

PA (GEST) GENSET.

PI Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.
 XX DR

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX
 XX Claim 1; SEQ ID 16793; 71pp + CD-ROM; English.

CC The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors.

XX Sequence 328 BP; 60 A; 112 C; 97 G; 56 T; 3 other;

Query Match 85.6%; Score 15.4; DB 21; Length 328;
 Best Local Similarity 94.1%; Pred. No. 4.8e+02;
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 gcacagcagcgagcagaag 18
 |||||
 Db 258 GAACGACGCGGACGAGACT 242

RESULT 14
 AAC03564

ID AAC03564 standard; cDNA; 341 BP.

AC AAC03564;

DT 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 3562.

KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping; ss.

OS Homo sapiens.

PN EP1033401-A2.

PD 06-SEP-2000.

PF 21-FEB-2000; 2000EP-0200610.

PR 26-FEB-1999; 99US-0122487.

PA (GEST) GENSET.

PI Dumas Milne Edwards J, Duclert A, Giordano J;

DR WPI; 2000-500381/45.
 DR P-PSDB; AAC03558.

XX

XX

XX

XX

XX

XX The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. An ORF has been identified within the
 CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs

CC derived from 30 different tissues. EST sequences usually correspond
 CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
 CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
 CC well suited for isolating cDNA sequences derived from the 5' ends of
 CC mRNAs and even in those cases where longer cDNA sequences have been
 CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
 CC mRNAs with intact 5' ends and can therefore be used to obtain full length
 CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
 CC gene therapy and chromosome mapping procedures. They are used to obtain
 CC upstream regulatory sequences and to design expression and secretion
 CC vectors.

SQ Sequence 341 BP; 68 A; 90 C; 115 G; 66 T; 2 other;

Query Match

Best Local Similarity 85.6%; Score 15.4; DB 21; Length 341;
 Pred. No. 4.8e+02;

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 gcagcagcgagcagaagt 18
 |||
 Db 216 gcagcagcgagcagaagt 232

RESULT 15

AAC79295
 ID AAC79295 standard; cDNA; 422 BP.

AC AAC79295;

DT 05-FEB-2001 (first entry)

DE Human lung tumour-specific cDNA #248.

KM Lung tumour protein; lung cancer; cytostatic; vaccine; ss.

OS Homo sapiens.

PN WO200060077-A2.

PD 12-OCT-2000.

PF 30-MAR-2000; 2000MO-US08560.

PR 02-APR-1999; 990S-0285323.

PR 09-AUG-1999; 990S-0370838.

PR 30-DEC-1999; 990S-0476235.

PR 03-MAR-2000; 2000US-0518809.

PA (CORI-) CORIXA CORP.

PI Reed SG, Lodes MJ, Mohamath R, Secrlist H;

DR WPI; 2000-638466/61.

PT Novel lung tumor polypeptides and polynucleotides, useful for
 detecting, monitoring or treating cancer, especially lung cancer -

PS Claim 3; Page 227; 243pp: English.

CC The present sequence is given in a specification relating to compounds
 CC for therapy and diagnosis of lung cancer. Polypeptides comprising at
 CC least an immunogenic part of a lung tumour protein are disclosed.
 CC The polypeptides are useful for inhibiting the development of cancer,
 CC especially lung cancer. Samples of T cells expressing the polypeptides
 CC may be used to inhibit the development of cancer. The polypeptides are
 CC also useful for detecting and monitoring the progression of cancer,
 CC especially lung cancer.

SQ Sequence 422 BP; 103 A; 113 C; 163 G; 42 T; 1 other;

Query Match 85.6%; Score 15.4; DB 21; Length 422;

Best Local Similarity 94.1%; Pred. No. 4.9e+02;
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 gcagcagcgagcagaagt 18
 . |||
 Db 166 gcagcagcgagcagaagt 182

Search completed: September 21, 2001, 02:42:38
 Job time: 11454 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2001, 23:27:52 ; Search time 4309.39 Seconds
(without alignments)
39.484 Million cell updates/sec

Title: US-09-138-735-8

Perfect score: 18
Sequence: 1 tgcagcagcgcgagaagt 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 segs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

```
EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*
44: gb_est44:*
45: gb_est45:*
46: gb_est46:*
47: gb_est47:*
48: gb_est48:*
49: gb_est49:*
50: gb_est50:*
51: gb_est51:*
52: gb_est52:*
53: gb_est53:*
54: gb_est54:*
55: gb_est55:*
56: gb_est56:*
57: gb_est57:*
58: gb_est58:*
59: gb_est59:*
60: gb_est60:*
61: gb_est61:*
62: gb_est62:*
63: gb_est63:*
64: gb_est64:*
65: gb_est65:*
66: gb_est66:*
67: gb_est67:*
68: gb_est68:*
69: gb_est69:*
70: gb_est70:*
71: gb_est71:*
72: gb_est72:*
73: gb_est73:*
74: gb_est74:*
75: gb_est75:*
76: gb_est76:*
77: gb_est77:*
78: gb_est78:*
79: gb_est79:*
80: gb_est80:*
81: gb_est81:*
82: gb_est82:*
83: gb_est83:*
84: gb_est84:*
85: gb_est85:*
86: gb_est86:*
87: gb_est87:*
88: gb_est88:*
89: gb_est89:*
90: gb_est90:*
91: gb_est91:*
92: gb_est92:*
93: gb_est93:*
94: gb_est94:*
95: gb_est95:*
96: gb_est96:*
97: gb_est97:*
98: gb_est98:*
99: gb_est99:*
100: gb_est100:*
101: gb_est101:*
102: gb_est102:*
103: gb_est103:*
104: gb_est104:*
105: gb_est105:*
106: gb_est106:*
107: gb_est107:*
108: gb_est108:*
109: gb_est109:*
110: gb_est110:*
111: gb_est111:*
112: gb_est112:*
113: gb_est113:*
114: gb_est114:*
115: gb_est115:*
116: gb_est116:*
```

117: gb_est148:*
118: gb_est149:*
119: gb_est150:*
120: gb_est151:*
121: gb_est152:*
122: gb_est153:*
123: gb_est154:*
124: gb_est155:*
125: gb_est156:*
126: gb_est157:*
127: gb_est158:*
128: gb_est159:*
129: gb_est160:*
130: gb_est161:*
131: gb_est162:*
132: gb_est163:*
133: gb_est164:*
134: gb_est165:*
135: gb_est166:*
136: gb_est167:*
137: gb_est168:*
138: gb_est169:*
139: gb_est170:*
140: gb_est171:*
141: gb_est172:*
142: gb_est173:*
143: gb_est174:*
144: gb_est175:*
145: gb_est176:*
146: gb_est177:*
147: gb_est178:*
148: gb_est179:*
149: gb_est180:*
150: gb_est181:*
151: gb_est182:*
152: gb_est183:*
153: gb_est184:*
154: gb_est185:*
155: gb_est186:*
156: gb_est187:*
157: gb_est188:*
158: gb_est189:*
159: gb_est190:*
160: gb_est191:*
161: gb_est192:*
162: gb_est193:*
163: gb_est194:*
164: gb_est195:*
165: gb_est196:*
166: gb_est197:*
167: gb_est198:*
168: gb_est199:*
169: gb_est200:*
170: gb_est201:*
171: gb_est202:*
172: gb_est203:*
173: gb_est204:*
174: gb_est205:*
175: gb_est206:*
176: gb_est207:*
177: gb_est208:*
178: gb_est209:*
179: gb_est210:*
180: gb_est211:*
181: gb_est212:*
182: gb_est213:*
183: gb_est214:*
184: gb_est215:*
185: gb_est216:*
186: gb_est217:*
187: gb_est218:*
188: gb_est219:*
189: gb_est220:*

190: gb_est110:*
191: gb_est111:*
192: gb_hic:*
193: em_gss_fun:*
194: em_gss_hum1:*
195: em_gss_hum2:*
196: em_gss_hum3:*
197: em_gss_hum4:*
198: em_gss_hum5:*
199: em_gss_hum6:*
200: em_gss_hum7:*
201: em_gss_hum8:*
202: em_gss_hum9:*
203: em_gss_inv1:*
204: em_gss_inv2:*
205: em_gss_inv3:*
206: em_gss_inv4:*
207: em_gss_inv5:*
208: em_gss_inv6:*
209: em_gss_inv7:*
210: em_gss_inv8:*
211: em_gss_inv9:*
212: em_gss_inv10:*
213: em_gss_inv11:*
214: em_gss_inv12:*
215: em_gss_inv13:*
216: em_gss_inv14:*
217: em_gss_inv15:*
218: em_gss_inv16:*
219: em_gss_inv17:*
220: em_gss_inv18:*
221: em_gss_inv19:*
222: em_gss_inv20:*
223: em_gss_inv21:*
224: em_gss_inv22:*
225: em_gss_inv23:*
226: em_gss_inv24:*
227: em_gss_inv25:*
228: em_gss_inv26:*
229: em_gss_inv27:*
230: em_gss_inv28:*
231: em_gss_inv29:*
232: em_gss_inv30:*
233: em_gss_inv31:*
234: em_gss_inv32:*
235: em_gss_inv33:*
236: em_gss_inv34:*
237: em_gss_inv35:*
238: em_gss_inv36:*
239: em_gss_inv37:*
240: em_gss_inv38:*
241: em_gss_inv39:*
242: em_gss_inv40:*
243: em_gss_inv41:*
244: em_gss_inv42:*
245: em_gss_inv43:*
246: em_gss_inv44:*
247: em_gss_inv45:*
248: em_gss_inv46:*
249: em_gss_inv47:*
250: em_gss_inv48:*
251: em_gss_inv49:*
252: em_gss_inv50:*
253: em_gss_inv51:*
254: em_gss_inv52:*
255: em_gss_inv53:*
256: em_gss_inv54:*
257: em_gss_inv55:*
258: em_gss_inv56:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	17	94.4	524	138	BE665288	BE665288 153467 MA
2	17	94.4	563	138	BE665781	BE665781 154938 MA
3	17	94.4	658	238	AZ086647	AZ086647 RPT-23-2
4	17	94.4	725	248	AZ070723	AZ070723 RPT-23-2
5	16.4	91.1	201	192	AK019382	AK019382 Mus muscu
6	16.4	91.1	203	25	AV158967	AV158967 AV158967
7	16.4	91.1	208	161	BB591907	BB591907 BB591907
8	16.4	91.1	220	161	BB595622	BB595622 BB595622
9	16.4	91.1	240	161	BB563512	BB563512 BB563512
10	16.4	91.1	307	165	BE244644	BE244644 TCBAPE2E00
11	16.4	91.1	337	23	AI642272	AI642272 v076a01.y
12	16.4	91.1	358	105	AL362372	AL362372 AL362372
13	16.4	91.1	421	138	BE651324	BE651324 UI-M-BH3-
14	16.4	91.1	421	232	AO680611	AO680611 HS_5496.A
15	16.4	91.1	424	144	BF116881	BF116881 uy2012.y
16	16.4	91.1	425	15	AI035972	AI035972 u050c11.r
17	16.4	91.1	442	13	AA867357	AA867357 vx97c08.r
18	16.4	91.1	443	154	BC516264	BC516264 Ectesped56
19	16.4	91.1	445	230	AO589778	AO589778 HS_2117.B
20	16.4	91.1	459	154	BG466312	BG466312 Ectesped51
21	16.4	91.1	471	10	AA638912	AA638912 v076a01.r
22	16.4	91.1	475	117	AW530300	AW530300 UI-R-BU0-
23	16.4	91.1	484	114	AW322968	AW322968 u054f09.y
24	16.4	91.1	489	239	AZ183201	AZ183201 SP_0191.A
25	16.4	91.1	502	113	AW230793	AW230793 u067h04.y
26	16.4	91.1	505	114	AW320029	AW320029 un13e11.x
27	16.4	91.1	528	222	FR0034786	FR0034786 Fugu rubr
28	16.4	91.1	531	139	BE744774	BE744774 601573161
29	16.4	91.1	532	165	BE290024	BE290024 601089630
30	16.4	91.1	554	240	AZ220432	AZ220432 Sheared D
31	16.4	91.1	558	239	AZ154969	AZ154969 SP_0040.A
32	16.4	91.1	569	231	AO647766	AO647766 RPT193-DP
33	16.4	91.1	577	166	BE304046	BE304046 601085519
34	16.4	91.1	577	251	AZ866126	AZ866126 2M0176E01
35	16.4	91.1	588	107	AU018956	AU018956 AU018956
36	16.4	91.1	593	222	FR0034779	FR0034779 Fugu rubr
37	16.4	91.1	596	250	AZ859556	AZ859556 2M0165012
38	16.4	91.1	615	122	AW918438	AW918438 EST349742
39	16.4	91.1	623	258	TA106601P	TA106601P T. brucei
40	16.4	91.1	628	108	AU132503	AU132503 AU132503
41	16.4	91.1	636	149	BF491885	BF491885 AT28649.5
42	16.4	91.1	654	192	AK003238	AK003238 Mus muscu
43	16.4	91.1	655	149	BF486916	BF486916 AT21486.5
44	16.4	91.1	665	236	AO946302	AO946302 Sheared D
45	16.4	91.1	679	149	BF486412	BF486412 AT20711.5

ALIGNMENTS

RESULT	1	ALIGNMENTS
LOCUS	BE665288	524 bp mRNA
DEFINITION	153467 MARC 4BOV Bos taurus CDNA 5', mRNA sequence.	08-SEP-2000
ACCESSION	BE665288	
VERSION	BE665288.1	GI:10025706
KEYWORDS	EST.	
SOURCE	COW.	
ORGANISM	Bos taurus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovine; Bos.	
AUTHORS	Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Laegreid ,W.W. and Keeler,J.W.	
TITLE	Design and use of four pooled tissue normalized CDNA libraries for	

JOURNAL COMMENT

EST discovery in cattle
Unpublished (2000)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390

Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt-trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -minmatch 12 options.

PCR primers
FORWARD: AGGAACAGCATGACCAT
BACKWARD: GTTTTCCAGTCACGAC
Plate: 69 row: D column: 6
Seq primer: ATTAGGTGACATATAG.
Location/Qualifiers

FEATURES

source
1..524
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 4BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site 1: XbaI; Site 2: XhoI; Library made from pooled tissue from day 20 and day 40 embryos."
BASE COUNT 106 a 145 c 153 g 120 t
ORIGIN

Query Match 94.4%; Score 17; DB 138; Length 524;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgcagcagcgagcgaag 17
|||||
Db 286 TGCAGCAGCGGAGAG 302

RESULT	2
LOCUS	BE665781 563 bp mRNA
DEFINITION	154938 MARC 4BOV Bos taurus CDNA 5', mRNA sequence.
ACCESSION	BE665781
VERSION	BE665781.1 GI:10026372
KEYWORDS	EST.
SOURCE	COW.
ORGANISM	Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovine; Bos.

1 (bases 1 to 563)

Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Laegreid ,W.W. and Keeler,J.W.

Design and use of four pooled tissue normalized CDNA libraries for

JOURNAL COMMENT

Unpublished (2000)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt-trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -minmatch 12 options.

PCR primers

FORWARD: AGGAACAGCATGACCAT
BACKWARD: GTTTTCCAGTCACGAC
Plate: 69 row: D column: 7
Seq primer: ATTAGGTGACATATAG.
Location/Qualifiers

source 1. .563
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_1lb="MARC 4BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: PCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from day 20 and day 40
embryos."
BASE COUNT 116 a 158 c 157 g 131 t 1 others
ORIGIN

Query Match 94.4%; Score 17; DB 138; Length 563;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgcagcagcgccagaag 17
|||||

Db 286 TGCAGCAGCGCCAGAG 302

RESULT 3
LOCUS AZ086647 658 bp DNA GSS 08-MAY-2000
DEFINITION RPCI-23-26F10.TV RPCI-23 Mus musculus genomic clone RPCI-23-26F10,
DNA sequence.
ACCESSION AZ086647
VERSION AZ086647.1 GI:7728381
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 658)
Zhao, S., Niernan, W., Feldblyum, T., Malek, J., Shatsman, S., Akınret,
B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.
and Fraser, C. M.
Mouse BAC End Sequences from library RPCI-23
Unpublished (1999)
Other GSSs: RPCI-23-26F10.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieter@ed.jong.med.bufileo.edu). Clones may be purchased from
BACPAC Resources (<http://bacpac.med.bufileo.edu/orderingframe.htm>)
or from Resea ch Genetics (<http://www.reschgenetics.com>). BAC end page:
http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html
Plate: 26 row: F column: 10
Seq primer: SP6
Class: BAC ends.

FEATURES
source Location/Qualifiers
1. .658
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-26F10"
/clone_1lb="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBACe3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 181 a 168 c 150 g 159 t
ORIGIN

Query Match 94.4%; Score 17; DB 238; Length 658;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgcagcagcgccagaag 17
|||||

Db 532 TGCAGCAGCGCCAGAG 548

RESULT 4
LOCUS AZ705723 725 bp DNA GSS 24-JAN-2001
DEFINITION RPCI-23-234M21.TV RPCI-23 Mus musculus genomic clone RPCI-23-234M21
, DNA sequence.
ACCESSION AZ705723
VERSION AZ705723.1 GI:12432498
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 725)
Zhao, S., Niernan, W., Feldblyum, T., Malek, J., Shatsman, S., Akınret,
B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.
and Fraser, C. M.
Mouse BAC End Sequences from library RPCI-23
Unpublished (1999)
Other GSSs: RPCI-23-234M21.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pdejong@emil.cho.org). Clones may be purchased from BACPAC
Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end
page: http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html
Plate: 234 row: M column: 21
Seq primer: 17
Class: BAC ends.

FEATURES
source Location/Qualifiers
1. .725
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-234M21"
/clone_1lb="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBACe3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 208 a 174 c 167 g 176 t
ORIGIN

Query Match 94.4%; Score 17; DB 248; Length 725;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgcagcagcgccagaag 17
|||||

Db	548	TGCACGACGGCAGGAAG	564
----	-----	-------------------	-----

RESULT	5		
LOCUS	AKO19382/c		
DEFINITION	Mus musculus 12 days embryo head cDNA, RIKEN full-length enriched library, clone:3010020L09, full insert sequence.	HTC	08-FEB-2001
ACCESSION	AKO19382		
VERSION	AKO19382.1		
KEYWORDS	GI:12859557		
SOURCE	CAP trapper. Mus musculus (strain:C57Bl/6J) 12 days embryo head cDNA to mRNA, clone:libRIKEN full-length enriched mouse cDNA library clone:3010020L09.		
ORGANISM	Mus musculus		
REFERENCE	Eumakryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (sites)		
TITLE	Carninci, P. and Hayashizaki, Y.		
JOURNAL	High-efficiency full-length cDNA cloning		
REFERENCE	Methods Enzymol. 303, 19-44 (1999)		
AUTHORS	2 (sites)		
TITLE	Carninci, P., Shibata, K., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.		
JOURNAL	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes		
REFERENCE	Genome Res. 10 (10), 1617-1630 (2000)		
MEDLINE	20499374		
REFERENCE	3 (sites)		
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M., Kikuchi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiyagi, K., Fujiwake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsuhira, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Okazaki, Y., Muramatsu, M., Inoue, Y. and Hayashizaki, Y.		
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer		
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)		
MEDLINE	20530913		
REFERENCE	4 (sites)		
AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and FANTOM Consortium.		
TITLE	Functional annotation of a full-length mouse cDNA collection		
JOURNAL	Nature 409, 685-690 (2001)		
REFERENCE	5 (bases 1 to 201)		
AUTHORS	Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Arkawa, T., Carninci, P., Fukuda, S., Fukunishi, Y., Fununo, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hiraoka, T., Horii, F., Imotani, K., Ishii, Y., Itoh, M., Ikawa, S., Kato, H., Kawaji, J., Koike, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sekai, C., Sekai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.		
TITLE	Direct Submission		
JOURNAL	Submitted (18-AUG-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)		
COMMENT	Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.		
	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5'-GAGACGAGAAGACATCAAGACTCTTTTTTTTTTTTTTTAA 3'], cDNA was		

prepared by using trehalase thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization for Rct = 10.0 and subtraction to Rct = 50.0. Second strand cDNA was prepared with the primer adapter of sequence 15',
GAGACGAGAGTTCGAGTTAAATTAATTAATGCCCCCCCCCC 3'. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
Host: SOLR.

FEATURES
source
Location/Qualifiers
1..201
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/db_xref="MGD:MGI:1899127"
/db_xref="MGD:MGI:1914915"
/clone="3010020L09"
/tissue="type="head"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="12 days embryo"
13..>201
/note="putative"
/codon_start=1
/protein_id="BAB31691.1"
/db_xref="GI:12859558"
/translation="MASLLOSERLYLVQGEKKRAFLSQLYFCRYSELRSLCEVSH
EVDSHYPCSELEMPASAA"
BASE COUNT 41 a 59 c 54 g 47 t
ORIGIN

Query Match 91.1%; Score 16.4; DB 192; Length 201;
Best Local Similarity 94.4%; Pred. No. 1.9e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

0Y 1 tgcacgacgacgacgaagt 18
||||| ||||||| |||||
Db 112 tgcactacgcgcgacgaagt 95

RESULT 6
AV158967/c
LOCUS
DEFINITION
AV158967 Mus musculus head C57BL/6J 12-day embryo Mus musculus cDNA
clone 3010020L09, mRNA sequence.
ACCESSION
AV158967
VERSION
AV158967.1 GI:5402602
KEYWORDS
EST.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Ciniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 203)
Carninci, P., Shibata, K., Ozawa, Y., Kono, H., Itoh, M., Aizawa, K.,
Akehita, S., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T., Hara
A., Hayatsu, N., Horii, F., Ishikawa, T., Itoh, M., Izawa, M., Kawai, J.,
Kikuchi, N., Kojima, Y., Matsuyama, T., Mitsuuma, H., Oda, H., Owa, C.,
Sato, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara
Y., Suzuki, H., Suzuki, H., Tateo, M., Tomaru, Y., Tomiwa, N.,
Watanabe, S., Yagame, M., Yamamura, T., Yokota, T., Yoshino, M.,
Miyamatsu, M., Okazaki, Y. and Hayashizaki, Y.
RIKEN Mouse ESTs
Unpublished (1999)
Contact: Chile Owa
Genome Science Laboratory
RIKEN
3-1-1 Koyada, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-9145
Fax: 81-298-36-9098
Email: genome-resetc.riken.go.jp
Thermotabilization and thermoactivation of therolabile enzymes by
trehalose and its application for the synthesis of full length cDNA
(Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
Transcriptional sequencing: A method for DNA sequencing using RNA

Contact: Yoshihide Hayashizaki
Genome Exploration Research Group, Life Science Tsukuba Center,
Genome Science Laboratory
The Institute of Physical and Chemical Research (RIKEN), Genomic
Sciences Center
3-1-1 Koyada, Tsukuba, Ibaraki 305-0074, Japan
Tel.: +81-298-36-9013
Fax: +81-298-36-9018
Email: genome-res@rtc.riken.go.jp,
URL: <http://genome.rtc.riken.go.jp/>
Carinichi, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoke, S., Sasakawa,
N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Thermostabilization and thermoactivation of thermolabile enzymes by
trehalose and its application for the syntheses of full length
cDNA. *Proc. Natl. Acad. Sci. U.S.A.* 95 (2), 520-524 (1998)
Itoh, M., Kitanaka, T., Akiyama, J., Shibata, K., Iwama, M., Kawai, J.,
Tomaru, Y., Carinichi, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki,
Y. and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation
system. *Genome Res.* 9 (5), 463-470 (1999)
Carinichi, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. *Methods Enzymol.* 303,
19-44 (1999)

1 (bases 1 to 220)
 Aizawa, K., Akahira, S., Akimura, T., Arai, A., Arakawa, T., Carninci, P.,
 Hanagaki, T., Hayatsu, N., Hirooka, T., Hirozane, T., Hotoyama, Y.,
 Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kawai, J., Kojima, Y., Konno,
 H., Kusakabe, M., Matsuyama, T., Miyazaki, A., Nakamura, M., Nishik, K.,
 Nomura, K., Nunazaki, R., Okazaki, Y., Okido, T., Owa, C., Sakai, C.,
 Sakai, K., Sasaki, D., Sato, K., Shibata, K., Shibata, Y., Shingawa, A.,
 Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Tanaka,
 T., Toya, T., Watanishi, A., Yamamura, T., Yasunishi, A., Yoshida, K.,
 Yoshiaki, A., Muramatsu, M. and Hayashizaki, Y.
 RIKEN Mouse ESTs (Aizawa, K. et al. 2000)
 Unpublished (2000)
 Contact: Yoshinobu Hayashizaki
 Genome Exploration Research Group, Life Science Tsukuba Center,
 The Institute of Physical and Chemical Research (RIKEN), Genomic
 Sciences Center
 3-3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
 Tel: +81-298-36-9013
 Fax: +81-298-36-9098
 Email: genome-research@riken.go.jp,
 URL: http://genome.rtc.riken.go.jp/

ACCESSION BE244644 GI:9096386
 VERSION BE244644.1
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 307)
 AUTHORS Wei,Y., Tsang,Y.T.M., Mei,G., Ku,J.M., Ali-Osman Jf.,F.R., Muzny,D., Bouck,J., Gibbs,R.A. and Margolin,J.F.
 TITLE Pediatric Leukemia cDNA Sequencing Project
 JOURNAL Unpublished (2000)
 COMMENT Contact: Dr. Judith F. Margolin
 Texas Children's Cancer Center and Human Genome Sequencing Center
 at Baylor College of Medicine
 1102 Bates, MC3-3320 Houston, TX 77030, USA
 Tel: 832-824-4536
 Fax: 832-825-4038
 Email: clones@tccc.org
 Seq primer: M13 primer.
 FEATURES
 source
 Location/Qualifiers
 1..307
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="TCBAP0065"
 /clone_lib="Pediatric pre-B cell acute lymphoblastic leukemia Baylor-HGSC project-TCBA"
 /sex="male"
 /tissue_type="leukopheresis"
 /cell_type="pre-B cell"
 /dev_stage="pediatric 2 years"
 /lab_host="DH10B"
 /note="Vector: lambda pSB: Site_1: BamHI; Site_2: EcoRI; First strand cDNA was primed with an anchored XhoI-colligo(dT) primer [5'GGAAGCAGTCGACGGCCGACGAGGAG(T)VN 3'; V-A,C,G; N-A,C,G,T] and then dgtailed. Second strand was primed with a BamHI-dc primer [5'AGAGAGCTCGATCCGCGCGCCCAATGAATAT(C) 3']. Double-stranded cDNA was then digested with BamHI and XhoI and directionally cloned into the BamHI and SalI sites of lambda pSB vector. Library went through one round of normalization. Library was constructed by Wei Yu"
 BASE COUNT 38 a 112 c 107 g 49 t 1 others
 ORIGIN
 Query Match 91.1%; Score 16.4; DB 165; Length 307;
 Best local Similarity 94.4%; Pred. No. 2e+03;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 tgcagcagcgagcagaagt 18
 ||||| ||||| ||||| |||||
 Db 232 TGCAGCGCGCGCAGAGT 215
 RESULT 11
 A1642272 337 bp mRNA EST 29-APR-1999
 A1642272/c LOCUS
 DEFINITION clone IMAGE:1037832 5', mRNA sequence.
 ACCESSION A1642272
 VERSION A1642272.1 GI:4720747
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 337)
 AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wyllie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shih,T., Jackson,Y., Cardenas,M., McCann,R.,

TITLE Waterston,R. and Wilson,R.
 JOURNAL The WashU-NCI Mouse EST Project 1999
 COMMENT Unpublished (1999)
 Contact: Maira M/WashU-NCI Mouse EST Project 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.
 MGI:577856
 This read is a RESEQUENCE of a previously sequenced mouse clone
 This read has been verified (found to hit its original self in the correct orientation)
 Seq primer: -40RP
 High quality sequence stop: 336.
 FEATURES
 source
 Location/Qualifiers
 1..337
 /organism="Mus musculus"
 /strain="B6D2 F1/J"
 /db_xref="taxon:10090"
 /clone="IMAGE:1037832"
 /clone_lib="Knowles Solter mouse blastocyst B1"
 /tissue_type="blastocyst"
 /dev_stage="embryo (pre-implantation)"
 /lab_host="DH10B"
 /note="Organ: embryo; Vector: pSPORT; Site_1: NotI; Site_2: SalI; Cloned unidirectionally from mRNA prepared from 800 blastocysts. Primer: SalI(dT): 5'-CGGTGACCGTCGACCGTGTGTGTGT-3'. cDNAs were cloned into the NotI/SalI sites of a pSPORT vector (Life Technologies). Two different size selections: B1 (larger inserts) and B3."
 BASE COUNT 54 a 92 c 117 g 72 t 2 others
 ORIGIN
 Query Match 91.1%; Score 16.4; DB 23; Length 337;
 Best local Similarity 94.4%; Pred. No. 2e+03;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 tgcagcagcgagcagaagt 18
 ||||| ||||| ||||| |||||
 Db 104 TGCAGTACGCGCAGAGT 87
 RESULT 12
 AL362372/c 358 bp mRNA EST 04-AUG-2000
 AL362372 LOCUS
 DEFINITION AL362372 ICRFP 522 and 523 Mus musculus cDNA clone W9322B43 5', mRNA sequence.
 ACCESSION AL362372
 VERSION AL362372.1 GI:9694940
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 358)
 AUTHORS Eickhoff,H., Schuchhardt,J., Ivanov,I., Meier-Ewert,S., O'Brien,J., Malik,A., Tandon,N., Wolski,E., Rohls,E., Nyarsik,L., Reinhardt,R., Nietfeld,W. and Lehnach,H.
 TITLE Tissue gene expression analysis using arrayed normalized cDNA libraries
 JOURNAL Genome Res. (2000) In press
 COMMENT Contact: MPIWG
 Abt. Lehnach
 Max Planck Institut fuer Molekulare Genetik
 Ihnestrasse 73, Berlin, 14195 Germany
 The cDNA libraries ICRFP 522 and 523 were normalized with oligonucleotide fingerprinting, resulting in a unique subset of 5376 cDNA clones.

FEATURES
source

Location/Qualifiers
1. .358
/organism="Mus musculus"
/strain="Black 6"
/db_xref="taxon:10090"
/clone="W9322B43"
/clone_lib="ICRF 522 and 523"
/tissue_type="embryo"
/dev_stage="9 and 12 pc embryo"
BASE COUNT 63 a 88 c 114 g 67 t 26 others
ORIGIN

Query Match 91.1%; Score 16.4; DB 105; Length 358;
Best Local Similarity 94.4%; Pred. No. 2e+03; 1; Indels 0; Gaps 0;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 tgcagcagcgagcagaagt 18
||||| |||||||||
Db 120 TGCAGTAGCGCAGAGACT 103

RESULT 13
BE651324 418 bp mRNA EST 06-SEP-2000
LOCUS BE651324/c
DEFINITION UI-M-BH3-asy-c-12-0-UI-r1 NIH BMAP_M.S4 Mus musculus cDNA clone
ACCESSION BE651324
VERSION BE651324.1 GI:9977148
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 418)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)

JOURNAL MEDLINE
COMMENT 97044477
Contact: Chln, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mstremail.nih.gov
CDNA Library Preparation: M.B. Soares Lab Clone Distribution:
Researchers may obtain BMAP CDNA clones from RESEARCH GENETICS. It
should be noted that Bento Soares is generating a small number of
additional specialized non-redundant arrays of BMAP cDNAs whose
availability will be considered under appropriate and limited
collaborative arrangements
Seq primer: M13 Reverse.
Location/Qualifiers
1. .418
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-BH3-asy-c-12-0-UI"
/clone_lib="NIH_BMAP_M.S4"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The
NIH_BMAP_M.S4 library is a subcloned library of a series,
ultimately derived from a mixture of individually tagged
normalized libraries from ten regions of the mouse brain
(cerebellum, brain stems, olfactory bulbs, hypothalamus,
cortex, amygdala, basal ganglia, pineal gland, striatum,
hippocampus) after a series of subtractions to reduce the
representation of cDNAs from which ESTs had already been
generated. The following serially subtracted libraries

FEATURES
source

Location/Qualifiers
1. .421
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=1072 Col=1 Row=C"
/clone_lib="RPCT-11 Human Male BAC Library"
/sex="male"

BASE COUNT 110 a 112 c 100 g 96 t
ORIGIN

Query Match 91.1%; Score 16.4; DB 138; Length 418;
Best Local Similarity 94.4%; Pred. No. 2e+03; 1; Indels 0; Gaps 0;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 tgcagcagcgagcagaagt 18
||||| |||||||||
Db 92 TGCAGTAGCGCAGAGACT 75

RESULT 14
AO680611 421 bp DNA GSS 25-JUN-1999
LOCUS AO680611
DEFINITION HS_5496.AL.B01.SP6E RPCT-11 Human Male BAC Library Homo sapiens
genomic clone Plate=1072 Col=1 Row=C, DNA sequence.
ACCESSION AO680611
VERSION AO680611.1 GI:5229415
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 421)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCT-11. For BAC
library availability, please contact Pieter de Jong
(pieterdejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Resear h Genetics (info@resgen.com). BAC end Web Server:
<http://www.htsc.washington.edu>
Plate: 1072 row: C column: 1
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 421.
Location/Qualifiers
1. .421
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=1072 Col=1 Row=C"
/clone_lib="RPCT-11 Human Male BAC Library"
/sex="male"

FEATURES
source

Location/Qualifiers
1. .421
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=1072 Col=1 Row=C"
/clone_lib="RPCT-11 Human Male BAC Library"
/sex="male"

/note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACE3.6 vector at EcoRI sites"
BASE COUNT 133 a 76 c 69 g 141 t 2 others
ORIGIN

Db 107 TGCAGTAGCGGAGAGT 90
||||| |||||||||

Search completed: September 20, 2001, 23:28:03
Job time: 25543 sec

Query Match 91.1%; Score 16.4; DB 232; Length 421;
Best local Similarity 94.4%; Pred. No. 2e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 tcgacgacgagcagaagt 18
||||| |||||
Db 90 TGCAGCAGCGGTAGAGT 107

RESULT 15
BF116881 424 bp mRNA EST 29-DEC-2000
LOCUS uy92d12.y1 NCI-CGAP Mam5 Mus musculus cDNA clone IMAGE:3667031 5'
DEFINITION similar to TR:Q9QXP8 Q9QXP8 P62 DYNACTIN SUBUNIT. [1] ;, mRNA
sequence.

ACCESSION BF116881
VERSION BF116881.1 GI:10986357
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 424)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
image.llnl.gov/image/html/resources.shtml

MGI:1427799
Seq primer: -40RP from Gibco
High quality sequence stop: 350.
Location/Qualifiers

FEATURES
source 1..424
/organism="Mus musculus"
/strain="C57/B6"
/db_xref="taxon:10090"
/clone_1b="IMAGE:3667031"
/clone_1lb="NCI-CGAP_Mam5"
/tissue_type="tumor, gross tissue"
/dev_stage="7 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Robin Humphreys,
NIH"

BASE COUNT 111 a 115 c 102 g 96 t
ORIGIN

Query Match 91.1%; Score 16.4; DB 144; Length 424;
Best local Similarity 94.4%; Pred. No. 2e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 tcgacgacgagcagaagt 18

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 21, 2001, 02:16:29 : Search time 164.23 Seconds
(without alignments)
20.749 Million cell updates/sec

Title: US-09-138-735-8

Perfect score: 18

Sequence: 1 tgcagcagcgagcagaagt 18

Scoring table: IDENTITY_NUC

Searched: 324599 seqs, 94655562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA:*

- 1: /cgn1_7/ptodata/1/ina/5A.COMB.seq:*
- 2: /cgn1_7/ptodata/1/ina/5B.COMB.seq:*
- 3: /cgn1_7/ptodata/1/ina/6A.COMB.seq:*
- 4: /cgn1_7/ptodata/1/ina/6B.COMB.seq:*
- 5: /cgn1_7/ptodata/1/ina/PCtUS.COMB.seq:*
- 6: /cgn1_7/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18	100.0	18	US-08-480-917-8	Sequence 8, Appli
2	18	100.0	3402	US-08-480-917-1	Sequence 1, Appli
3	18	100.0	11873	US-08-970-269A-32	Sequence 32, Appli
4	18	100.0	11878	US-08-970-269A-31	Sequence 31, Appli
5	18	100.0	11883	US-08-970-269A-28	Sequence 28, Appli
6	16.4	91.1	36519	US-08-923-137-2	Sequence 2, Appli
7	15.4	85.6	39	US-09-110-959A-13	Sequence 13, Appli
8	15.4	85.6	1276	US-09-177-325-2	Sequence 2, Appli
9	15.4	85.6	2000	US-07-923-739-1	Sequence 1, Appli
10	15.4	85.6	2214	US-08-864-038A-1	Sequence 1, Appli
11	15.4	85.6	3331	US-08-864-038A-2	Sequence 2, Appli
12	15.4	85.6	3331	US-08-864-038A-4	Sequence 4, Appli
13	15.4	85.6	5197	US-08-131-365B-53	Sequence 53, Appli
14	15.4	85.6	5197	US-08-668-123-53	Sequence 53, Appli
15	15.4	85.3	1869	US-08-934-494-1	Sequence 1, Appli
16	15.4	83.3	1869	US-09-143-068-1	Sequence 1, Appli
17	14.8	82.2	38	US-08-470-369-24	Sequence 24, Appli
18	14.8	82.2	38	5304637-24	Patent No. 5304637
19	14.8	82.2	1033	US-08-414-657D-19	Sequence 19, Appli
20	14.8	82.2	1851	US-08-414-657D-20	Sequence 20, Appli
21	14.8	82.2	2232	US-08-179-481-97	Sequence 97, Appli
22	14.8	82.2	2464	US-07-863-169A-4	Sequence 4, Appli
23	14.8	82.2	2464	US-08-429-964-4	Sequence 4, Appli
24	14.8	82.2	2464	US-07-935-087-4	Sequence 4, Appli
25	14.8	82.2	2464	PCR-US93-08062-4	Sequence 4, Appli
26	14.8	82.2	2603	US-08-179-481-1	Sequence 1, Appli
27	14.8	82.2	2716	US-08-647-484-1	Sequence 1, Appli

28	14.8	82.2	2716	1	US-08-647-484-3	Sequence 3, Appli
29	14.8	82.2	2716	1	US-08-647-481-1	Sequence 1, Appli
30	14.8	82.2	2716	1	US-08-647-481-3	Sequence 3, Appli
31	14.8	82.2	2716	1	US-08-430-033A-1	Sequence 1, Appli
32	14.8	82.2	2716	1	US-08-430-033A-3	Sequence 3, Appli
33	14.8	82.2	2716	5	PCR-US96-05792-1	Sequence 1, Appli
34	14.8	82.2	2716	5	PCR-US96-05792-3	Sequence 3, Appli
35	14.8	82.2	5183	3	US-09-039-555B-18	Sequence 18, Appli
36	14.8	82.2	6942	2	US-08-460-309-3	Sequence 3, Appli
37	14.8	82.2	6942	2	US-08-125-077-3	Sequence 3, Appli
38	14.4	80.0	19	1	US-08-410-540-5	Sequence 5, Appli
39	14.4	80.0	110	1	US-07-668-648-11	Sequence 11, Appli
40	14.4	80.0	110	2	US-08-429-998-11	Sequence 11, Appli
41	14.4	80.0	110	2	US-08-431-333-11	Sequence 11, Appli
42	14.4	80.0	110	5	PCR-US91-02321-11	Sequence 11, Appli
43	14.4	80.0	306	2	US-08-634-797-14	Sequence 14, Appli
44	14.4	80.0	707	2	US-08-850-910A-40	Sequence 40, Appli
45	14.4	80.0	971	3	US-09-248-335-65	Sequence 65, Appli

ALIGNMENTS

RESULT 1
US-08-480-917-8
; Sequence 8, Application US/08480917
; Patent No. 5820864
; GENERAL INFORMATION:
; APPLICANT: PARANHOS-BACCALA, Glaucia
; APPLICANT: LESENECHAL, Mylene
; APPLICANT: JOLIVET, Michel
; TITLE OF INVENTION: NEW TRYPAOSOMA CRUZI ANTIGEN, AND GENE
; TITLE OF INVENTION: ENCODING THE LATTER, THEIR APPLICATION TO THE DETECTION OF
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Oliff & Berridge
; STREET: 700 South Washington Street, Suite 300
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,917
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Berridge, William P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: MPB 36400
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6400
; TELEFAX: 703-836-2787
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-480-917-8

Query Match 100.0%; Score 18; DB 1; Length 18;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 tgcagcagcgagcagaagt 18

DB 1 TGCAGCAGCGCAGAACT 18

RESULT 2

US-08-480-917-1/c
; Sequence 1, Application US/08480917
; Patent No. 5820864
; GENERAL INFORMATION:
; APPLICANT: PARANHOS-BACCALA, Glaucia
; APPLICANT: LESENECHAL, Mylene
; APPLICANT: JOIIVET, Michel
; TITLE OF INVENTION: NEW TRIPANOSOMA CRUZI ANTIGEN, AND GENE
; TITLE OF INVENTION: ENCODING THE LATTER; THEIR APPLICATION TO THE DETECTION OF
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oliff & Berridge
; STREET: 700 South Washington Street, Suite 300
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,917
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Berridge, William P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPP 36400
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6400
; TELEFAX: 703-836-2787
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3402 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-480-917-1

Query Match 100.0%; Score 18; DB 1; Length 3402;

Best Local Similarity 100.0%; Pred. No. 11;

Matches 18; Conservative 0; Indels 0; Gaps 0;

QY 1 tgcagcagcgcgcaagct 18

DB 1459 TGCAGCAGCGCAGAACT 1442

RESULT 3

US-08-970-269A-32
; Sequence 32, Application US/08970269A
; Patent No. 5976803
; GENERAL INFORMATION:
; APPLICANT: Katheryn Meek
; TITLE OF INVENTION: Genetic Test For Equine Severe
; TITLE OF INVENTION: Combined Immunodeficiency Disease
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dr. Benjamin A. Adler
; STREET: 8011 Candle Lane
; CITY: Houston
; STATE: Texas
; COUNTRY: USA

ZIP: 77071

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple
; OPERATING SYSTEM: Macintosh
; SOFTWARE: Microsoft Word for Macintosh

CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/970,269A
; FILING DATE: No. 5976803ember 14, 1997
; CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
; NAME: Adler Ph.D., Benjamin A.

REGISTRATION NUMBER: 35,423

REFERENCE/DOCKET NUMBER: D5860

TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713-777-2321

TELEFAX: 713-777-6908

INFORMATION FOR SEQ ID NO: 32:

SEQUENCE CHARACTERISTICS:

LENGTH: 11873 bp

TYPE: nucleic acid

STRANDEDNESS: double stranded

TOPOLOGY: linear

MOLECULE TYPE:

DESCRIPTION: other nucleic acid

HYPOTHETICAL: no

ANTI-SENSE: no

ORIGINAL SOURCE:

FEATURE:

US-08-970-269A-32

Query Match 100.0%; Score 18; DB 2; Length 11873;

Best Local Similarity 100.0%; Pred. No. 12;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgcagcagcgcgcaagct 18

DB 6429 TGCAGCAGCGCAGAACT 6446

RESULT 4

US-08-970-269A-31
; Sequence 31, Application US/08970269A
; Patent No. 5976803
; GENERAL INFORMATION:
; APPLICANT: Katheryn Meek
; TITLE OF INVENTION: Genetic Test For Equine Severe
; TITLE OF INVENTION: Combined Immunodeficiency Disease
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dr. Benjamin A. Adler
; STREET: 8011 Candle Lane
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple
; OPERATING SYSTEM: Macintosh
; SOFTWARE: Microsoft Word for Macintosh
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/970,269A
; FILING DATE: No. 5976803ember 14, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Adler Ph.D., Benjamin A.
; REGISTRATION NUMBER: 35,423
; REFERENCE/DOCKET NUMBER: D5860
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713-777-2321
; TELEFAX: 713-777-6908

INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 11878 bp
TYPE: nucleic acid
STRANDEDNESS: double stranded
TOPOLOGY: linear
MOLECULE TYPE:
DESCRIPTION: other nucleic acid
HYPOTHETICAL: no
ANTI-SENSE: no
ORIGINAL SOURCE:
FEATURE:
US-08-970-269A-31

Query Match 100.0%; Score 18; DB 2; Length 11878;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 tgcagcagcgagcagaagt 18
|||||
Db 6429 TGCACGACGCGCAGAGT 6446

RESULT 5
US-08-970-269A-28
Sequence 28, Application US/08970269A
Patent No. 5976803
GENERAL INFORMATION:
APPLICANT: Kathryn Meek
TITLE OF INVENTION: Genetic Test For Equine Severe
TITLE OF INVENTION: Combined Immunodeficiency Disease
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dr. Benjamin A. Adler
STREET: 8011 Candle Lane
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77071
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple
OPERATING SYSTEM: Macintosh
SOFTWARE: Microsoft Word for Macintosh
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/970,269A
FILING DATE: No. 5976803ember 14, 1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Adler Ph.D., Benjamin A.
REGISTRATION NUMBER: 35,423
REFERENCE/DOCKET NUMBER: D5860
TELECOMMUNICATION INFORMATION:
TELEPHONE: 713-777-2321
TELEFAX: 713-777-6908
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 11883 bp
TYPE: nucleic acid
STRANDEDNESS: double stranded
TOPOLOGY: linear
MOLECULE TYPE:
DESCRIPTION: other nucleic acid
HYPOTHETICAL: no
ANTI-SENSE: no
ORIGINAL SOURCE:
US-08-970-269A-28

Query Match 100.0%; Score 18; DB 2; Length 11883;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 tgcagcagcgagcagaagt 18
|||||
Db 6432 TGCACGACGCGCAGAGT 6449

RESULT 6
US-08-923-137-2
Sequence 2, Application US/08923137
Patent No. 6083716
GENERAL INFORMATION:
APPLICANT: Wilson, James M.
APPLICANT: Farina, Steven F.
APPLICANT: Fisher, Krishna J.
TITLE OF INVENTION: Chimpanzee Adenovirus Vectors
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Cntr., P.O. Box 457
CITY: Spring House
STATE: Pennsylvania
COUNTRY: United States of America
ZIP: 19477
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/923,137
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/024,700
FILING DATE: 06-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: GNPVN.021CIP1USA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 36519 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: CDNA
US-08-923-137-2

Query Match 91.1%; Score 16.4; DB 3; Length 36519;
Best Local Similarity 94.4%; Pred. No. 63;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 tgcagcagcgagcagaagt 18
|||||
Db 31662 TGCACGACGCGCAGAGT 31679

RESULT 7
US-09-110-959A-13
Sequence 13, Application US/09110959A
Patent No. 6268197
GENERAL INFORMATION:
APPLICANT: Schuelein, Martin
APPLICANT: Outtrup, Helle
APPLICANT: Jorgensen, Per Lina
APPLICANT: Bjornvad, Mads Eskelund
TITLE OF INVENTION: Alkaline Xyloglucanase
FILE REFERENCE: 5206.200-US
CURRENT APPLICATION NUMBER: US/09/110,959A

```

; CURRENT FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 0822/97
; PRIOR FILING DATE: 1997-07-07
; PRIOR APPLICATION NUMBER: 1213/97
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/054,039
; PRIOR FILING DATE: 1997-07-28
; PRIOR APPLICATION NUMBER: 60/063,694
; PRIOR FILING DATE: 1997-10-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 39
; TYPE: DNA
; ORGANISM: Bacillus sp.
US-09-110-959A-13

Query Match      85.6%; Score 15.4; DB 4; Length 39;
Best Local Similarity 94.1%; Pred. No. 98;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 tgcagcgcgcgcgcagaag 17
Db      6 tgcagcgcgcgcgcagaag 22

RESULT      8
US-09-177-325-2
; Sequence 2, Application US/09177325B
; Patent No. 6214983
; GENERAL INFORMATION:
; APPLICANT: Tureci, Ozlem
; APPLICANT: Sakin, Ugur
; APPLICANT: Pfreundschuh, Michael
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Which Encode SCP Proteins, A
; FILE REFERENCE: Theresof
; CURRENT APPLICATION NUMBER: US/09/177,325B
; CURRENT FILING DATE: 1998-10-22
; NUMBER OF SEQ ID NOS: 15
; SEQ ID NO 2
; LENGTH: 1276
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-177-325-2

Query Match      85.6%; Score 15.4; DB 4; Length 1276;
Best Local Similarity 94.1%; Pred. No. 1.3e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 gccagcgcgcgcgcagaagt 18
Db      711 gccagcgcgcgcgcagaagt 727

RESULT      9
US-07-923-739-1
; Sequence 1, Application US/07923739
; Patent No. 5401835
; GENERAL INFORMATION:
; APPLICANT: Chishtli, Athar H.
; TITLE OF INVENTION: Human Erythroid p55 and Methods of Use
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Harness, Dickey and Pierce
; STREET: 5445 Corporate Drive
; CITY: Troy
; STATE: Michigan
; COUNTRY: US
; ZIP: 48098
```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/923,739
; FILING DATE: 19920731
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lewak, Anna M
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (313) 641-1600
; TELEFAX: (313) 641-0270
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2000 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: Single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 115..1512
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /product="protein"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /note="Polyadenylation consensus sequence is at
; OTHER INFORMATION: residues 165-233 represent the SH-3 motif."
; PUBLICATION INFORMATION:
; AUTHORS: Andirabi, Khurshid
; AUTHORS: Rana, Ajay
; AUTHORS: Keeler, Marilyn
; AUTHORS: Maalouf, George
; AUTHORS: Bruns, Gail
; AUTHORS: Chishtli, Athar
; TITLE: Human erythroid p55: Homolog of Drosophila
; TITLE: tumor suppressor factor is highly conserved
; TITLE: X-linked gene product with guanylate kinase
; TITLE: activity
; JOURNAL: J. Biol. Chem.
; DATE: 1992
; PUBLICATION INFORMATION:
; AUTHORS: Ruff, Paul
; AUTHORS: Speicher, David W.
; AUTHORS: Husain-Chishtli, A.
; TITLE: Molecular identification of a major
; TITLE: palmitoylated erythrocyte membrane protein
; TITLE: containing the src homology 3 motif
; JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
; VOLUME: 88
; PAGES: 6595-6599
; DATE: August-1991
; RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 2000
; PUBLICATION INFORMATION:
; AUTHORS: Husain-Chishtli, Athar
; AUTHORS: Faquin, William
; AUTHORS: Wu, Chi-Chih
; AUTHORS: Branton, Daniel
; TITLE: Purification of Erythrocyte of Dematin
; TITLE: (Protein 4.9) Reveals an Endogenous Protein Kinase
; TITLE: That Modulates Actin-bundling Activity
; JOURNAL: J. Biol. Chem.
; VOLUME: 264
; ISSUE: 15
; PAGES: 8985-8991
; DATE: 5-25-1989
US-07-923-739-1

Query Match      85.6%; Score 15.4; DB 1; Length 2000;
Best Local Similarity 94.1%; Pred. No. 1.3e+02;
```

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 tgcagcagcgagcaag 17
|||||

Db 1349 TGCAGCAGCTGCAGCAG 1365

RESULT 10

US-08-864-038A-1/c

; Sequence 1, Application US/08864038A

; Patent No. 6001592

; GENERAL INFORMATION:

; APPLICANT: Kunio NAKASHIMA et al.

; TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, VECTOR

; TITLE OF INVENTION: CONTAINING SAID CDNA, HOST CELLS TRANSFORMED WITH SAID

; TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING

; TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBODY

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: 812-5 Hirano

; STREET: Ishinden

; CITY: Tsu-city

; STATE: Mie-prefecture

; COUNTRY: JAPAN

; ZIP: 514-01

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: Microsoft Windows 95

; SOFTWARE: Word Perfect 6.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/864,038A

; FILING DATE: May 28, 1997

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 8-184459

; FILING DATE: 15-July-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: C. Bruce Hamburg

; REGISTRATION NUMBER: 22,389

; REFERENCE/DOCKET NUMBER: F-5610

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212)986-2340

; TELEFAX: (212)953-7733

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2214

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; ORIGINAL SOURCE:

; ORGANISM: Pinctada fucata

; CELL TYPE: mantle epithelial cell

; US-08-864-038A-1

Query Match 85.6%; Score 15.4; DB 3; Length 2214;
Best Local Similarity 94.1%; Pred. No. 1.4e+02;

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 tgcagcagcgagcaag 17
|||||

Db 1767 TGCAGCAGCGCAGCAG 1751

RESULT 11

US-08-864-038A-2/c

; Sequence 2, Application US/08864038A

; Patent No. 6001592

; GENERAL INFORMATION:

; APPLICANT: Kunio NAKASHIMA et al.

; TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, VECTOR

; TITLE OF INVENTION: CONTAINING SAID CDNA, HOST CELLS TRANSFORMED WITH SAID

; TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING

; TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBOD

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: 812-5 Hirano

; STREET: Ishinden

; CITY: Tsu-city

; STATE: Mie-prefecture

; COUNTRY: JAPAN

; ZIP: 514-01

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: Microsoft Windows 95

; SOFTWARE: Word Perfect 6.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/864,038A

; FILING DATE: May 28, 1997

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 8-184459

; FILING DATE: 15-July-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: C. Bruce Hamburg

; REGISTRATION NUMBER: 22,389

; REFERENCE/DOCKET NUMBER: F-5610

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212)986-2340

; TELEFAX: (212)953-7733

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 3331

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: CDNA to mRNA

; ORIGINAL SOURCE:

; ORGANISM: Pinctada fucata

; CELL TYPE: mantle epithelial cell

; FEATURE: mRNA

; LOCATION: from 1 to 3331

; IDENTIFICATION METHOD: E (by experiment)

; US-08-864-038A-2

Query Match 85.6%; Score 15.4; DB 3; Length 3331;
Best Local Similarity 94.1%; Pred. No. 1.4e+02;

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 tgcagcagcgagcaag 17
|||||

Db 1816 TGCAGCAGCGCAGCAG 1800

RESULT 12

US-08-864-038A-4/c

; Sequence 4, Application US/08864038A

; Patent No. 6001592

; GENERAL INFORMATION:

; APPLICANT: Kunio NAKASHIMA et al.

; TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, VECTOR

; TITLE OF INVENTION: CONTAINING SAID CDNA, HOST CELLS TRANSFORMED WITH SAID

; TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING

; TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBOD

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: 812-5 Hirano

; STREET: Ishinden

; CITY: Tsu-city

; STATE: Mie-prefecture

; COUNTRY: JAPAN

; ZIP: 514-01

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Word Perfect 6.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/864,038A
FILING DATE: May 28, 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 8-184459
FILING DATE: 15-July-1996
ATTORNEY/AGENT INFORMATION:
NAME: C. Bruce Hamburg
REGISTRATION NUMBER: 22,389
REFERENCE/DOCKET NUMBER: F-5610
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)986-2340
TELEFAX: (212)953-7733
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 3331
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
ORGANISM: Pinctada fucata
CELL TYPE: mantle epithelial cell
FEATURE:
NAME/KEY: CDS
LOCATION: from 50 to 2263
IDENTIFICATION METHOD: P (by similarity to some other pattern)
US-08-864-038A-4

Query Match 85.6%; Score 15.4; DB 3; Length 3331;
Best Local Similarity 94.1%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 tgcagcagcgagcagaag 17
|||||
Db 1816 TGCAGCAGCGGCGAGCAG 1800

RESULT 13
US-08-131-365B-53
Sequence 53, Application US/08131365B
Patent No. 5527690
GENERAL INFORMATION:
APPLICANT: Brown, Michael S.
APPLICANT: Briggs, Michael R.
APPLICANT: Wang, Xiaodong
APPLICANT: Goldstein, Joseph L.
TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING
TITLE OF INVENTION: TO STEROL REGULATORY ELEMENT BINDING
TITLE OF INVENTION: PROTEINS
NUMBER OF SEQUENCES: 64
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: U.S.A.
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/131,365B
FILING DATE: 01-OCT-1993
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Parker, David L.
REGISTRATION NUMBER: 32,165
REFERENCE/DOCKET NUMBER: UTSD:372/PAR
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 53:
SEQUENCE CHARACTERISTICS:
LENGTH: 5197 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
US-08-131-365B-53

Query Match 85.6%; Score 15.4; DB 1; Length 5197;
Best Local Similarity 94.1%; Pred. No. 1.5e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 gcagcagcgagcagaagt 18
|||||
Db 302 GCAGCAGCGGCGAGCAGT 318

RESULT 14
US-08-668-123-53
Sequence 53, Application US/08668123
Patent No. 5891631
GENERAL INFORMATION:
APPLICANT: Brown, Michael S.
APPLICANT: Briggs, Michael R.
APPLICANT: Wang, Xiaodong
APPLICANT: Goldstein, Joseph L.
TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING
TITLE OF INVENTION: TO STEROL REGULATORY ELEMENT BINDING
TITLE OF INVENTION: PROTEINS
NUMBER OF SEQUENCES: 64
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: U.S.A.
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/668,123
FILING DATE: 14-JUN-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/131,365
FILING DATE: 01-OCT-1993
ATTORNEY/AGENT INFORMATION:
NAME: Parker, David L.
REGISTRATION NUMBER: 32,165
REFERENCE/DOCKET NUMBER: UTSD:372/PAR
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 53:
SEQUENCE CHARACTERISTICS:
LENGTH: 5197 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid

DESCRIPTION: /desc = "DNA"
US-08-668-123-53

Query Match 85.6%; Score 15.4; DB 2; Length 5197;
Best Local Similarity 94.1%; Pred. No. 1.5e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 gcagcagcgagcagaagt 18
|||||
Db 302 GCAGCAGCGCAGCAGT 318

RESULT 15

US-08-934-494-1
Sequence 1, Application US/08934494
Patent No. 6030831
GENERAL INFORMATION:
APPLICANT: Gurney, Austin
TITLE OF INVENTION: Tie Ligands
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 MB floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/934,494
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Dregler, Ginger R.
REGISTRATION NUMBER: 33,055
REFERENCE/DOCKET NUMBER: P1078
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-3216
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1869 base pairs
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
US-08-934-494-1

Query Match 83.3%; Score 15; DB 3; Length 1869;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 cagcagcgagcagaag 17
|||||
Db 1268 CAGCAGCGCAGCAGAG 1282

Search completed: September 21, 2001, 02:16:34
Job time: 28227 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 21, 2001, 02:19:42 : Search time 6407.28 Seconds
(without alignments)
41.513 Million cell updates/sec

Title: US-09-138-735-8
Perfect score: 18
Sequence: 1 tgcacagcgcgcagcaagt 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1415048 seqs, 7388405095 residues

Total number of hits satisfying chosen parameters: 28310096

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Pending_Patents_NA_Main:*
1: /cgnl_7/ptodata/1/pna/PCRUS.COMB.seq:*
2: /cgnl_7/ptodata/1/pna/US06.COMB.seq:*
3: /cgnl_7/ptodata/1/pna/US07.COMB.seq:*
4: /cgnl_7/ptodata/1/pna/US08.COMB.seq:*
5: /cgnl_7/ptodata/1/pna/US081.COMB.seq:*
6: /cgnl_7/ptodata/1/pna/US082.COMB.seq:*
7: /cgnl_7/ptodata/1/pna/US083.COMB.seq:*
8: /cgnl_7/ptodata/1/pna/US084.COMB.seq:*
9: /cgnl_7/ptodata/1/pna/US085.COMB.seq:*
10: /cgnl_7/ptodata/1/pna/US086.COMB.seq:*
11: /cgnl_7/ptodata/1/pna/US087.COMB.seq:*
12: /cgnl_7/ptodata/1/pna/US088.COMB.seq:*
13: /cgnl_7/ptodata/1/pna/US089.COMB.seq:*
14: /cgnl_7/ptodata/1/pna/US090.COMB.seq:*
15: /cgnl_7/ptodata/1/pna/US091.COMB.seq:*
16: /cgnl_7/ptodata/1/pna/US092.COMB.seq:*
17: /cgnl_7/ptodata/1/pna/US093.COMB.seq:*
18: /cgnl_7/ptodata/1/pna/US094.COMB.seq:*
19: /cgnl_7/ptodata/1/pna/US095A.COMB.seq:*
20: /cgnl_7/ptodata/1/pna/US095B.COMB.seq:*
21: /cgnl_7/ptodata/1/pna/US095C.COMB.seq:*
22: /cgnl_7/ptodata/1/pna/US095D.COMB.seq:*
23: /cgnl_7/ptodata/1/pna/US096A.COMB.seq:*
24: /cgnl_7/ptodata/1/pna/US096B.COMB.seq:*
25: /cgnl_7/ptodata/1/pna/US096C.COMB.seq:*
26: /cgnl_7/ptodata/1/pna/US096D.COMB.seq:*
27: /cgnl_7/ptodata/1/pna/US096E.COMB.seq:*
28: /cgnl_7/ptodata/1/pna/US097A.COMB.seq:*
29: /cgnl_7/ptodata/1/pna/US097B.COMB.seq:*
30: /cgnl_7/ptodata/1/pna/US097C.COMB.seq:*
31: /cgnl_7/ptodata/1/pna/US098.COMB.seq:*
32: /cgnl_7/ptodata/1/pna/US099.COMB.seq:*
33: /cgnl_7/ptodata/1/pna/US6001.COMB.seq:*
34: /cgnl_7/ptodata/1/pna/US6002.COMB.seq:*
35: /cgnl_7/ptodata/1/pna/US6003.COMB.seq:*
36: /cgnl_7/ptodata/1/pna/US6004.COMB.seq:*
37: /cgnl_7/ptodata/1/pna/US6005.COMB.seq:*
38: /cgnl_7/ptodata/1/pna/US6006.COMB.seq:*
39: /cgnl_7/ptodata/1/pna/US6007.COMB.seq:*
40: /cgnl_7/ptodata/1/pna/US6008.COMB.seq:*
41: /cgnl_7/ptodata/1/pna/US6009.COMB.seq:*
42: /cgnl_7/ptodata/1/pna/US6010.COMB.seq:*
43: /cgnl_7/ptodata/1/pna/US6011.COMB.seq:*

44: /cgnl_7/ptodata/1/pna/US6012.COMB.seq:*
45: /cgnl_7/ptodata/1/pna/US6013.COMB.seq:*
46: /cgnl_7/ptodata/1/pna/US6014.COMB.seq:*
47: /cgnl_7/ptodata/1/pna/US6015.COMB.seq:*
48: /cgnl_7/ptodata/1/pna/US6016.COMB.seq:*
49: /cgnl_7/ptodata/1/pna/US6017.COMB.seq:*
50: /cgnl_7/ptodata/1/pna/US6018.COMB.seq:*
51: /cgnl_7/ptodata/1/pna/US6019.COMB.seq:*
52: /cgnl_7/ptodata/1/pna/US6020.COMB.seq:*
53: /cgnl_7/ptodata/1/pna/US6021.COMB.seq:*
54: /cgnl_7/ptodata/1/pna/US6022.COMB.seq:*
55: /cgnl_7/ptodata/1/pna/US6023.COMB.seq:*
56: /cgnl_7/ptodata/1/pna/US6024.COMB.seq:*
57: /cgnl_7/ptodata/1/pna/US6025.COMB.seq:*
58: /cgnl_7/ptodata/1/pna/US6026.COMB.seq:*
59: /cgnl_7/ptodata/1/pna/US6027.COMB.seq:*
60: /cgnl_7/ptodata/1/pna/US6028.COMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB	ID	Description
1	18	100.0	18	13	US-08-988-242-4	Sequence 4, Appl1
2	18	100.0	18	15	US-09-138-736-8	Sequence 8, Appl1
3	18	100.0	328	1	PCT-US01-01329-161	Sequence 161, App
4	18	100.0	328	1	PCT-US01-01339-1027	Sequence 1027, App
5	18	100.0	444	30	US-09-770-175-4591	Sequence 4591, App
6	18	100.0	475	16	US-09-240-371-7431	Sequence 7431, App
7	18	100.0	484	23	US-09-253-972-5995	Sequence 5995, App
8	18	100.0	484	25	US-09-606-976-3048	Sequence 3048, App
9	18	100.0	484	25	US-09-644-867-3664	Sequence 3664, App
10	18	100.0	506	6	US-08-276-163A-13709	Sequence 13709, A
11	18	100.0	506	6	US-08-276-163B-13709	Sequence 13709, A
12	18	100.0	506	6	US-08-276-163D-13709	Sequence 13709, A
13	18	100.0	508	25	US-09-652-914-2822	Sequence 2822, App
14	18	100.0	940	23	US-09-606-976-5253	Sequence 5253, App
15	18	100.0	940	25	US-09-770-175-8299	Sequence 8299, App
16	18	100.0	1589	25	US-09-641-377-783	Sequence 783, App
17	18	100.0	1707	25	US-09-644-867-5948	Sequence 5948, App
18	18	100.0	1707	25	US-09-652-914-7995	Sequence 7995, App
19	18	100.0	3402	13	US-08-988-242-1	Sequence 1, Appl1
20	18	100.0	11873	18	US-09-407-562-32	Sequence 32, Appl1
21	18	100.0	11873	18	US-09-407-562-31	Sequence 31, Appl1
22	18	100.0	11883	18	US-09-407-562-28	Sequence 28, Appl1
23	18	100.0	11883	18	US-09-407-562-28	Sequence 28, Appl1
24	17	94.4	429	49	US-09-127-612-136	Sequence 136, App
25	17	94.4	718	25	US-09-654-617-126156	Sequence 126156, App
26	17	94.4	718	27	US-09-684-016-126156	Sequence 126156, App
27	17	94.4	2121	18	US-09-489-039A-5009	Sequence 5009, App
28	17	94.4	914740	20	US-09-534-859-19	Sequence 19, App
29	17	94.4	249	18	US-09-436-762A-26553	Sequence 26553, App
30	16.4	91.1	280	15	US-09-127-612-1967	Sequence 1967, App
31	16.4	91.1	280	15	US-09-540-208-69979	Sequence 69979, App
32	16.4	91.1	281	15	US-09-127-612-2550	Sequence 2550, App
33	16.4	91.1	281	21	US-09-540-208-68661	Sequence 68661, App
34	16.4	91.1	284	48	US-09-164-443-401	Sequence 401, App
35	16.4	91.1	307	20	US-09-535-897-16555	Sequence 16555, App
36	16.4	91.1	339	21	US-09-540-229-178511	Sequence 178511, App
37	16.4	91.1	402	17	US-09-333-335A-2478	Sequence 2478, App
38	16.4	91.1	402	25	US-09-654-617-56139	Sequence 56139, App
39	16.4	91.1	402	27	US-09-684-016-56139	Sequence 56139, App
40	16.4	91.1	412	56	US-09-248-798-473	Sequence 473, App
41	16.4	91.1	423	27	US-09-654-617-258006	Sequence 258006, App
42	16.4	91.1	423	27	US-09-684-016-258006	Sequence 258006, App
43	16.4	91.1	486	18	US-09-401-645-3287	Sequence 3287, App
44	16.4	91.1	486	25	US-09-644-870-269	Sequence 269, App
45	16.4	91.1	489	53	US-09-217-080-20722	Sequence 20722, App

ALIGNMENTS

```
RESULT 1
US-08-988-242-4
; Sequence 4, Application US/08988242
; GENERAL INFORMATION:
; APPLICANT: PARANHOS-BACCALA, GLAUCIA
; APPLICANT: LESENECHAL, MYLENE
; APPLICANT: JOLIVET, MICHEL
; APPLICANT: MANDRAND, BERNARD
; TITLE OF INVENTION: NEW TRIPANOSOMA CRUZI ANTIGEN, GENE
; TITLE OF INVENTION: ENCODING THEREFOR, AND METHODS OF DETECTING AND TREATING
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OLIFF & BERRIDGE, PLC
; STREET: P.O. BOX 19928
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/988,242
; FILING DATE: 10-DEC-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Berridge, William P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: MPB 36400A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6400
; TELEFAX: 703-836-2787
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-988-242-4

Query Match 100.0%; Score 18; DB 13; Length 18;
Best Local Similarity 100.0%; Pred. No. 7.2e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgcagcagcgcgagaagt 18
Db 1 TGCAGCAGCGCAGAGAAGT 18

RESULT 2
US-09-138-736-8
; Sequence 8, Application US/09138736
; GENERAL INFORMATION:
; APPLICANT: PARANHOS-BACCALA, GLAUCIA
; APPLICANT: LESENECHAL, MYLENE
; APPLICANT: JOLIVET, MICHEL
; TITLE OF INVENTION: NEW TRIPANOSOMA CRUZI ANTIGEN, AND GENE
; TITLE OF INVENTION: ENCODING THE LATTER, THEIR APPLICATION TO THE DETECTION OF
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OLIFF & Berridge
; STREET: 700 South Washington Street, Suite 300
; CITY: Alexandria
```

```
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/138,736
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/480,917
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Berridge, William P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: MPB 36400
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6400
; TELEFAX: 703-836-2787
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-138-736-8
```

```
Query Match 100.0%; Score 18; DB 15; Length 18;
Best Local Similarity 100.0%; Pred. No. 7.2e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 tgcagcagcgcgagaagt 18
Db 1 TGCAGCAGCGCAGAGAAGT 18
```

```
RESULT 3
PCT-US01-01329-161
; Sequence 161, Application PC/TUS0101329
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc., et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA120PCT
; CURRENT APPLICATION NUMBER: PCT/US01/01329
; CURRENT FILING DATE: 2001-01-17
; NUMBER OF SEQ ID NOS: 3506
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 161
; LENGTH: 328
; TYPE: DNA
; ORGANISM: Homo sapiens
PCT-US01-01329-161
```

```
Query Match 100.0%; Score 18; DB 1; Length 328;
Best Local Similarity 100.0%; Pred. No. 7.9e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 tgcagcagcgcgagaagt 18
Db 234 tgcagcagcgcgagaagt 251
```

```
RESULT 4
PCT-US01-01339-1027
; Sequence 1027, Application PC/TUS0101339
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc., et al.
```

```
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006PCT
; CURRENT APPLICATION NUMBER: PCT/US01/01339
; CURRENT FILING DATE: 2001-03-17
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1027
; LENGTH: 328
; TYPE: DNA
; ORGANISM: Homo sapiens
PCT-US01-01339-1027
```

```
Query Match          100.0%; Score 18; DB 1; Length 328;
Best Local Similarity 100.0%; Pred. No. 7.9e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1  tgcagcagcgcgagaagt 18
          |||
Db      234  tgcagcagcgcgagaagt 251
```

```
RESULT      5
US-09-770-175-4591
; Sequence 4591, Application US/09770175
; GENERAL INFORMATION:
; APPLICANT: Gearling, David P.
; APPLICANT: Holtzman, Douglas A.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; FILE REFERENCE: 1600.2058-001
; CURRENT APPLICATION NUMBER: US/09/770.175
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: US 60/178,874
; PRIOR FILING DATE: 2000-01-28
; NUMBER OF SEQ ID NOS: 8967
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4591
; LENGTH: 444
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(444)
; OTHER INFORMATION: n = A,T,C or G
US-09-770-175-4591
```

```
Query Match          100.0%; Score 18; DB 30; Length 444;
Best Local Similarity 100.0%; Pred. No. 8e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1  tgcagcagcgcgagaagt 18
          |||
Db      255  tgcagcagcgcgagaagt 272
```

```
RESULT      6
US-09-240-371-7431
; Sequence 7431, Application US/09240371
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-761
; CURRENT APPLICATION NUMBER: US/09/240.371
; CURRENT FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 10214
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7431
; LENGTH: 475
; TYPE: DNA
; ORGANISM: Homo sapiens
```

```
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(475)
; OTHER INFORMATION: n = A,T,C or G
US-09-240-371-7431
```

```
Query Match          100.0%; Score 18; DB 16; Length 475;
Best Local Similarity 100.0%; Pred. No. 8e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1  tgcagcagcgcgagaagt 18
          |||
Db      315  tgcagcagcgcgagaagt 332
```

```
RESULT      7
US-09-293-972-5995
; Sequence 5995, Application US/09293972
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-769
; CURRENT APPLICATION NUMBER: US/09/293.972
; CURRENT FILING DATE: 1999-04-15
; NUMBER OF SEQ ID NOS: 34258
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5995
; LENGTH: 475
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(475)
; OTHER INFORMATION: n = A,T,C or G
US-09-293-972-5995
```

```
Query Match          100.0%; Score 18; DB 16; Length 475;
Best Local Similarity 100.0%; Pred. No. 8e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1  tgcagcagcgcgagaagt 18
          |||
Db      315  tgcagcagcgcgagaagt 332
```

```
RESULT      8
US-09-606-776-3048
; Sequence 3048, Application US/09606776
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Myers, Paul
; APPLICANT: Gearling, David P.
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 1600.1129-001
; CURRENT APPLICATION NUMBER: US/09/606.776
; CURRENT FILING DATE: 2000-06-27
; PRIOR APPLICATION NUMBER: 60/141,578
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: 60/141,379
; PRIOR FILING DATE: 1999-06-28
; PRIOR APPLICATION NUMBER: 60/141,138
; PRIOR FILING DATE: 1999-06-28
; PRIOR APPLICATION NUMBER: 60/141,581
; PRIOR FILING DATE: 1999-06-29
; NUMBER OF SEQ ID NOS: 5415
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3048
; LENGTH: 484
```

TYPE: DNA
ORGANISM: Homo sapiens
US-09-606-776-3048

Query Match 100.0%; Score 18; DB 23; Length 484;
Best Local Similarity 100.0%; Pred. No. 8e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgcagcagcgagcagaagt 18
|||||

DB 443 tgcagcagcgagcagaagt 465

RESULT 9

US-09-644-867-3664
Sequence 3664, Application US/09644867
GENERAL INFORMATION:
APPLICANT: Holtzman, Douglas A.
APPLICANT: Donovan, Michael J.
APPLICANT: Leiby, Kevin R.
APPLICANT: Culpepper, Janice A.
TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
FILE REFERENCE: 1600.1173-001
CURRENT APPLICATION NUMBER: US/09/644,867
CURRENT FILING DATE: 2000-08-28
PRIOR APPLICATION NUMBER: 60/151,061
PRIOR FILING DATE: 1999-08-27
NUMBER OF SEQ ID NOS: 8090
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3664
LENGTH: 484
TYPE: DNA
ORGANISM: Homo sapiens
US-09-644-867-3664

Query Match 100.0%; Score 18; DB 25; Length 484;
Best Local Similarity 100.0%; Pred. No. 8e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgcagcagcgagcagaagt 18
|||||

DB 448 tgcagcagcgagcagaagt 465

RESULT 10

US-08-276-163A-13709
Sequence 13709, Application US/08276163A
GENERAL INFORMATION:
APPLICANT: Haseltine, William
APPLICANT: Fitzgerald, Lisa
APPLICANT: Adams, Mark
APPLICANT: Lee, Normal
APPLICANT: Fuldner, Rebecca
APPLICANT: Fleischmann, Robert
APPLICANT: Bult, Carol
APPLICANT: Blake, Judy
APPLICANT: White, Owen
APPLICANT: Clayton, Rebecca
APPLICANT: Pellegrino, Susan
TITLE OF INVENTION: Human Genes, Sequences, and Expression
NUMBER OF SEQUENCES: 15312
CORRESPONDENCE ADDRESS:
ADDRESSEE: Carella, Byrne, Bain, Gillfillan, Cecchi, Stewart, &
STREET: 6 Becker Farm Road
CITY: Roseland
STATE: New Jersey
COUNTRY: USA
ZIP: 07068

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage

COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 5.0
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/276,163A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: Oistein, Elliot
REGISTRATION NUMBER: 24,025
REFERENCE/DOCKET NUMBER: 325800-175
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201) 994-1700
TELEFAX: (201) 994-1744
INFORMATION FOR SEQ ID NO: 13709:
SEQUENCE CHARACTERISTICS:
LENGTH: 506 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-276-163A-13709

Query Match 100.0%; Score 18; DB 6; Length 506;
Best Local Similarity 100.0%; Pred. No. 8e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgcagcagcgagcagaagt 18
|||||

DB 226 tgcagcagcgagcagaagt 243

RESULT 11

US-08-276-163B-13709
Sequence 13709, Application US/08276163B
GENERAL INFORMATION:
APPLICANT: Adams, Mark D.
APPLICANT: Blake, Carol J.
APPLICANT: Blake, Judith A.
APPLICANT: Fitzgerald, Lisa
APPLICANT: Fleischmann, Robert D.
APPLICANT: Fraser, Claire M.
APPLICANT: Fuldner, Rebecca A.
APPLICANT: Gocayne, Jeannine D.
APPLICANT: Kelley, Jenny
APPLICANT: Kirkness, Ewen F.
APPLICANT: Sulten, Granger G., III
APPLICANT: Haseltine, William A.
APPLICANT: Rosen, Craig A.
APPLICANT: Ruben, Steve
APPLICANT: Dillon, Patrick J.
APPLICANT: Li, Haodong
TITLE OF INVENTION: Human Genes, Sequences and Expression Products
NUMBER OF SEQUENCES: 15314
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/276,163B
FILING DATE: JULY 15, 1994
CLASSIFICATION: 435

```

: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Michele M. Wales
: REGISTRATION NUMBER: P-43,975
: REFERENCE/DOCKET NUMBER: PO-14
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (301) 309-8504
: TELEFAX: (301) 309-8439
: INFORMATION FOR SEQ ID NO: 13709:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 506 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: US-08-276-163B-13709

Query Match          100.0%; Score 18; DB 6; Length 506;
Best Local Similarity 100.0%; Pred. No. 8e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 tgcagcagcgagcagaagt 18
        |||
Db       226 TGCAGCAGCGGAGAGT 243

RESULT 12
US-08-276-163D-13709
: Sequence 13709, Application US/08276163D
: GENERAL INFORMATION:
: APPLICANT: Adams, et. al.
: TITLE OR INVENTION: Human Genes, Sequences, and Expression Products
: FILE REFERENCE: P014
: CURRENT APPLICATION NUMBER: US/08/276,163D
: CURRENT FILING DATE: 1994-07-15
: NUMBER OF SEQ ID NOS: 15314
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 13709
: LENGTH: 506
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc feature
: LOCATION: (1)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (11)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (30)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (44)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (46)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (51)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (60)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (66)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (148)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (185)
```

```

: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (306)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (321)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (345)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (359)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (364)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (375)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (385)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (425)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (427)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (428)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (440)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (443)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (470)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (473)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (476)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (478)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (489)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (492)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (494)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (501)
: OTHER INFORMATION: n equals a,t,g, or c
: US-08-276-163D-13709

Query Match          100.0%; Score 18; DB 6; Length 506;
Best Local Similarity 100.0%; Pred. No. 8e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 tgcagcagcgagcagaagt 18
        |||
```

Db 226 tcgcagcagcgagcaagt 243

RESULT 13

Sequence 2822, Application US/09652914

GENERAL INFORMATION:

APPLICANT: Shyjan, Andrew W.

TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES

FILE REFERENCE: 1600.1193-001

CURRENT FILING DATE: 2000-08-31

PRIOR FILING DATE: 1999-08-31

NUMBER OF SEQ ID NOS: 9677

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 2822

LENGTH: 508

TYPE: DNA

ORGANISM: Homo sapiens

US-09-652-914-2822

Query Match 100.0%; Score 18; DB 25; Length 508;

Best Local Similarity 100.0%; Pred. No. 8e+02;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgcagcagcgagcaagt 18

Db 219 tcgcagcagcgagcaagt 236

RESULT 14

US-09-606-776-5253

Sequence 5253, Application US/09606776

GENERAL INFORMATION:

APPLICANT: Holtzman, Douglas A.

APPLICANT: Myers, Paul

APPLICANT: Geating, David P.

APPLICANT: Pan, Yang

TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES

FILE REFERENCE: 1600.1129-001

CURRENT FILING DATE: 2000-06-27

PRIOR FILING DATE: 1999-06-29

PRIOR FILING DATE: 1999-06-29

PRIOR FILING DATE: 1999-06-28

PRIOR FILING DATE: 1999-06-28

PRIOR FILING DATE: 1999-06-28

PRIOR FILING DATE: 1999-06-28

PRIOR FILING DATE: 1999-06-28

PRIOR FILING DATE: 1999-06-28

PRIOR FILING DATE: 1999-06-28

PRIOR FILING DATE: 1999-06-28

PRIOR FILING DATE: 1999-06-28

PRIOR FILING DATE: 1999-06-28

PRIOR FILING DATE: 1999-06-28

PRIOR FILING DATE: 1999-06-28

PRIOR FILING DATE: 1999-06-28

PRIOR FILING DATE: 1999-06-28

PRIOR FILING DATE: 1999-06-28

PRIOR FILING DATE: 1999-06-28

PRIOR FILING DATE: 1999-06-28

PRIOR FILING DATE: 1999-06-28

PRIOR FILING DATE: 1999-06-28

PRIOR FILING DATE: 1999-06-28

PRIOR FILING DATE: 1999-06-28

PRIOR FILING DATE: 1999-06-28

PRIOR FILING DATE: 1999-06-28

PRIOR FILING DATE: 1999-06-28

PRIOR FILING DATE: 1999-06-28

PRIOR FILING DATE: 1999-06-28

PRIOR FILING DATE: 1999-06-28

PRIOR FILING DATE: 1999-06-28

PRIOR FILING DATE: 1999-06-28

PRIOR FILING DATE: 1999-06-28

US-09-770-175-8299

Sequence 8299, Application US/09770175

GENERAL INFORMATION:

APPLICANT: Geating, David P.

APPLICANT: Holtzman, Douglas A.

TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES

FILE REFERENCE: 1600.2058-001

CURRENT FILING DATE: 2001-01-26

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

PRIOR FILING DATE: 2000-01-28

Search completed: September 21, 2001, 02:19:48
Job time: 30460 sec

Query Match 100.0%; Score 18; DB 30; Length 940;

Best Local Similarity 100.0%; Pred. No. 8.1e+02;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgcagcagcgagcaagt 18

Db 524 tcgcagcagcgagcaagt 541

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 21, 2001, 02:39:09 : Search time 1035.34 Seconds
(without alignments)
34.697 Million cell updates/sec

Title: US-09-138-735-8

Perfect score: 18

Sequence: 1 tgcagcagcgagcaagt 18

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapept 1.0

Searched: 1741792 seqs, 997871973 residues

Total number of hits satisfying chosen parameters: 3483584

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Pending_Patents_NA_New:*
1: /cgnl_7/ptodata/2/pna/PCT_NEW_COMB.seq:*
2: /cgnl_7/ptodata/2/pna/US06_NEW_COMB.seq:*
3: /cgnl_7/ptodata/2/pna/US07_NEW_COMB.seq:*
4: /cgnl_7/ptodata/2/pna/US08_NEW_COMB.seq:*
5: /cgnl_7/ptodata/2/pna/US09_NEW_COMB.seq:*
6: /cgnl_7/ptodata/2/pna/US09_NEW_COMB.seq1.*
7: /cgnl_7/ptodata/2/pna/US09_NEW_COMB.seq2.*
8: /cgnl_7/ptodata/2/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	18	100.0	18	6	US-09-138-735-8	Sequence 8, Appli
2	18	100.0	328	6	US-09-764-891-1027	Sequence 1027, Ap
3	18	100.0	506	7	US-09-840-145-13709	Sequence 13709, A
4	18	100.0	3402	6	US-09-138-735-1	Sequence 1, Appli
5	17	94.4	91470	6	US-09-803-736-19	Sequence 19, Appli
6	16.4	91.1	381	7	US-09-724-671-9810	Sequence 9810, Ap
7	16.4	91.1	506	6	US-09-804-730-4138	Sequence 4138, Ap
8	16.4	91.1	3826	7	US-09-927-091-3	Sequence 3, Appli
9	16.4	91.1	30676	7	US-09-927-091-8	Sequence 8, Appli
10	16.4	91.1	45845	7	US-09-927-091-6	Sequence 6, Appli
11	16.4	91.1	49744	7	US-09-927-091-4	Sequence 4, Appli
12	16	88.9	235	7	US-09-724-671-4868	Sequence 4868, Ap
13	16	88.9	326	7	US-09-724-671-18201	Sequence 18201, A
14	16	88.9	338	7	US-09-724-671-18461	Sequence 18461, A
15	16	88.9	363	7	US-09-724-671-18790	Sequence 18790, A
16	16	88.9	2268	1	PCT-US01-08631-19259	Sequence 19259, A
17	15.4	85.6	205	7	US-09-867-716-6478	Sequence 6478, Ap
18	15.4	85.6	215	8	US-60-253-654-23696	Sequence 23696, A
19	15.4	85.6	235	8	US-60-253-592-23696	Sequence 23696, A
20	15.4	85.6	235	7	US-09-850-147-1383	Sequence 1383, Ap
21	15.4	85.6	243	6	US-09-790-435-157	Sequence 157, App
22	15.4	85.6	256	7	US-09-880-107-541	Sequence 541, App
23	15.4	85.6	264	6	US-09-764-864-499	Sequence 499, App
24	15.4	85.6	266	8	US-60-253-456-735	Sequence 735, App
25	15.4	85.6	279	8	US-60-253-456-5135	Sequence 5135, App

C	26	15.4	85.6	317	8	US-60-253-456-23527	Sequence 23527, A
C	27	15.4	85.6	321	7	US-09-724-750-19995	Sequence 19995, A
C	28	15.4	85.6	332	6	US-09-758-472-1626	Sequence 1626, Ap
C	29	15.4	85.6	353	8	US-60-252-833-6389	Sequence 6389, Ap
C	30	15.4	85.6	359	7	US-09-867-716-17212	Sequence 17212, A
C	31	15.4	85.6	360	7	US-09-864-761-28274	Sequence 28274, A
C	32	15.4	85.6	378	7	US-09-850-118-114	Sequence 114, App
C	33	15.4	85.6	380	5	US-09-847-662-305	Sequence 305, App
C	34	15.4	85.6	389	6	US-09-790-435-2837	Sequence 2837, App
C	35	15.4	85.6	391	5	US-09-847-662-1609	Sequence 1609, Ap
C	36	15.4	85.6	393	8	US-60-252-833-42024	Sequence 42024, A
C	37	15.4	85.6	421	7	US-09-834-366-3070	Sequence 3070, Ap
C	38	15.4	85.6	430	6	US-09-790-435-6901	Sequence 6901, Ap
C	39	15.4	85.6	435	6	US-09-790-435-852	Sequence 852, App
C	40	15.4	85.6	440	6	US-09-906-555-12800	Sequence 12800, A
C	41	15.4	85.6	441	7	US-09-834-366-3865	Sequence 3865, A
C	42	15.4	85.6	449	8	US-60-253-652-17754	Sequence 17754, A
C	43	15.4	85.6	450	7	US-09-864-761-11696	Sequence 11696, A
C	44	15.4	85.6	456	6	US-09-909-629-34330	Sequence 34330, A
C	45	15.4	85.6	456	7	US-09-864-761-1296	Sequence 1296, Ap

ALIGNMENTS

RESULT 1
US-09-138-735-8
Sequence 8, Application US/09138735
GENERAL INFORMATION:
APPLICANT: PARAMOS-BACCALA, Glaucia
APPLICANT: LESENECHAL, Mylene
APPLICANT: JOUVET, Michel
TITLE OF INVENTION: TRYPAECOSOMA CRUZI ANTIGEN, GENE ENCODING THEREFOR AND METHODS
FILE REFERENCE: MPB 36400B
CURRENT APPLICATION NUMBER: US/09/138, 735
PRIOR FILING DATE: 1998-08-24
PRIOR APPLICATION NUMBER: US 08/480, 917
PRIOR FILING DATE: 1995-06-07
PRIOR APPLICATION NUMBER: FR 94/10132
PRIOR FILING DATE: 1994-08-12
NUMBER OF SEQ ID NOS: 13
SOFTWARE: PatentIn version 3.0
SEQ ID NO 8
LENGTH: 18
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: probe/primer
US-09-138-735-8

Query Match 100.0%; Score 18; DB 6; Length 18;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgcagcagcgagcaagt 18
|||||
Db 1 tgcagcagcgagcaagt 18

RESULT 2
US-09-764-891-1027
Sequence 1027, Application US/09764891
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC006
CURRENT APPLICATION NUMBER: US/09/764,891
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PAMM or file wrapper
NUMBER OF SEQ ID NOS: 10231
SOFTWARE: PatentIn Ver. 2.0

```

; SEQ ID NO 1027
; LENGTH: 328
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-1027

```

```

Query Match          100.0%; Score 18; DB 6; Length 328;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 tgcagcagcgagcagaagt 18
        |||
Db       234 tgcagcagcgagcagaagt 251

```

```

RESULT 3
US-09-840-145-13709
; Sequence 13709, Application US/09840145
; GENERAL INFORMATION:
; APPLICANT: Adams, et. al.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products
; FILE REFERENCE: POLICI
; CURRENT APPLICATION NUMBER: US/09/840,145
; PRIOR FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: 08/276,163
; NUMBER OF SEQ ID NOS: 15314
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 13709
; LENGTH: 506
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (11)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (30)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (44)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (46)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (51)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (60)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (66)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (148)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (185)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (306)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (321)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (345)
; OTHER INFORMATION: n equals a,t,g, or c

```

```

; NAME/KEY: misc_feature
; LOCATION: (359)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (364)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (375)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (385)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (425)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (427)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (428)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (430)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (440)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (443)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (470)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (473)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (476)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (478)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (489)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (492)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (494)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (501)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-840-145-13709

```

```

Query Match          100.0%; Score 18; DB 7; Length 506;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 tgcagcagcgagcagaagt 18
        |||
Db       226 tgcagcagcgagcagaagt 243

```

```

RESULT 4
US-09-138-735-1/c
; Sequence 1, Application US/09138735
; GENERAL INFORMATION:
; APPLICANT: PARANHOS-BACCALA, Glaucia
; APPLICANT: LESENECHAL, Mylene
; APPLICANT: JOLIVET, Michel

```

```

; TITLE OF INVENTION: TRYPAOSOMA CRUZI ANTIGEN, GENE ENCODING THEREFOR AND METHODS OF
; FILE REFERENCE: DETECTING AND TREATING CHAGAS DISEASE
; FILE REFERENCE: WPB 36400B
; CURRENT APPLICATION NUMBER: US/09/138,735
; CURRENT FILING DATE: 1998-08-24
; PRIOR APPLICATION NUMBER: US 08/480,917
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: FR 94/10132
; PRIOR FILING DATE: 1994-08-12
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 3402
; TYPE: DNA
; ORGANISM: Trypanosoma cruzi
US-09-138-735-1
```

```

Query Match          100.0%; Score 18; DB 6; Length 3402;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 tgcagcagcgcgagaagt 18
|||||
Db 1459 TGCACGACGCGCAGAGT 1442
```

```

RESULT 5
US-09-803-736-19
; Sequence 19, Application US/09803736
; GENERAL INFORMATION:
; APPLICANT: Bush, David F.
; APPLICANT: Levin, Irena M.
; APPLICANT: Norris, Susan R.
; APPLICANT: Rounsley, Steven D.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Plant Polymorphic Markers and Uses Thereof
; FILE REFERENCE: 38-10(15493)D
; CURRENT APPLICATION NUMBER: US/09/803,736
; CURRENT FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 09/534,859
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: Identified by Attorney Docket number 045983.0206CPUS01 38-10
; PRIOR FILING DATE: 2000-10-20
; NUMBER OF SEQ ID NOS: 1582
; SEQ ID NO 19
; LENGTH: 91470
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-803-736-19
```

```

Query Match          94.4%; Score 17; DB 6; Length 91470;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 tgcagcagcgcgagaagt 17
|||||
Db 42731 tgcagcagcgcgagaagt 42747
```

```

RESULT 6
US-09-724-671-9810/C
; Sequence 9810, Application US/09724671
; GENERAL INFORMATION:
; APPLICANT: Watson, James D
; APPLICANT: Murison, James G
; TITLE OF INVENTION: Polynucleotides, polypeptides expressed
; TITLE OF INVENTION: by the polynucleotides and methods for their use.
; FILE REFERENCE: 105002
; CURRENT APPLICATION NUMBER: US/09/724,671
; CURRENT FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: U.S. No. 60/171,678
```

```

; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 21907
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9810
; LENGTH: 381
; TYPE: DNA
; ORGANISM: Mouse
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: (1)...(381)
; OTHER INFORMATION: n = A,T,C or G
US-09-724-671-9810
```

```

Query Match          91.1%; Score 16.4; DB 7; Length 381;
Best Local Similarity 94.4%; Pred. No. 3.8e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
OY 1 tgcagcagcgcgagaagt 18
|||||
Db 106 TGCAGTACGCGCAGAGT 89
```

```

RESULT 7
US-09-804-730-4138
; Sequence 4138, Application US/09804730
; GENERAL INFORMATION:
; APPLICANT: Andersen, Scott E.
; APPLICANT: Hammond-Kosack, Kim
; APPLICANT: Masucci, James D.
; APPLICANT: Urban, Martin
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; FILE REFERENCE: 38-21(51838)B
; CURRENT APPLICATION NUMBER: US/09/804,730
; CURRENT FILING DATE: 2001-03-13
; PRIOR APPLICATION NUMBER: US 60/189,657
; PRIOR FILING DATE: 2000-03-15
; NUMBER OF SEQ ID NOS: 22828
; SEQ ID NO 4138
; LENGTH: 506
; TYPE: DNA
; ORGANISM: Trifolium aestivum
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)...(506)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: LIB3350-063-P1-K1-A3
US-09-804-730-4138
```

```

Query Match          91.1%; Score 16.4; DB 6; Length 506;
Best Local Similarity 94.4%; Pred. No. 3.8e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
OY 1 tgcagcagcgcgagaagt 18
|||||
Db 385 tgcagcagcgcgagaagt 402
```

```

RESULT 8
US-09-927-091-3/C
; Sequence 3, Application US/09927091
; GENERAL INFORMATION:
; APPLICANT: KILARY, ANN
; APPLICANT: LOTT, STEVE
; APPLICANT: CHANDLER, DAWN
; TITLE OF INVENTION: THE TUMOR SUPPRESSOR CAR-1
; FILE REFERENCE: UTSC:651US
; CURRENT APPLICATION NUMBER: US/09/927,091
; CURRENT FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 60/227,560
; PRIOR FILING DATE: 2000-08-23
```

; PRIOR APPLICATION NUMBER: 60/225,033
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 3
; LENGTH: 3826
; TYPE: DNA
; ORGANISM: Human
US-09-927-091-3

Query Match 91.1%; Score 16.4; DB 7; Length 3826;
Best Local Similarity 94.4%; Pred. No. 4.3e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 tgcagcagcgagcaagt 18
||||| |||||||
Db 706 TGCAGCGCGCGCAGACT 689

RESULT 9
US-09-927-091-8/c
; Sequence 8, Application US/09927091
; GENERAL INFORMATION:
; APPLICANT: KILLARY, ANN
; APPLICANT: LOTT, STEVE
; APPLICANT: CHANDLER, DAMN
; TITLE OF INVENTION: THE TUMOR SUPPRESSOR CAR-1
; FILE REFERENCE: UTSC:651US
; CURRENT APPLICATION NUMBER: US/09/927,091
; CURRENT FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 60/227,560
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: 60/225,033
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 8
; LENGTH: 30676
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: modified base
; LOCATION: (6671)..(30676)
; OTHER INFORMATION: n = A or C or G or T/U
US-09-927-091-8

Query Match 91.1%; Score 16.4; DB 7; Length 30676;
Best Local Similarity 94.4%; Pred. No. 4.8e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 tgcagcagcgagcaagt 18
||||| |||||||
Db 288 TGCAGCGCGCGCAGACT 271

RESULT 10
US-09-927-091-6/c
; Sequence 6, Application US/09927091
; GENERAL INFORMATION:
; APPLICANT: KILLARY, ANN
; APPLICANT: LOTT, STEVE
; APPLICANT: CHANDLER, DAMN
; TITLE OF INVENTION: THE TUMOR SUPPRESSOR CAR-1
; FILE REFERENCE: UTSC:651US
; CURRENT APPLICATION NUMBER: US/09/927,091
; CURRENT FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 60/227,560
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: 60/225,033
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 9

; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 6
; LENGTH: 45845
; TYPE: DNA
; ORGANISM: Human
US-09-927-091-6

Query Match 91.1%; Score 16.4; DB 7; Length 45845;
Best Local Similarity 94.4%; Pred. No. 4.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 tgcagcagcgagcaagt 18
||||| |||||||
Db 24868 TGCAGCGCGCGCAGACT 24851

RESULT 11
US-09-927-091-4/c
; Sequence 4, Application US/09927091
; GENERAL INFORMATION:
; APPLICANT: KILLARY, ANN
; APPLICANT: LOTT, STEVE
; APPLICANT: CHANDLER, DAMN
; TITLE OF INVENTION: THE TUMOR SUPPRESSOR CAR-1
; FILE REFERENCE: UTSC:651US
; CURRENT APPLICATION NUMBER: US/09/927,091
; CURRENT FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 60/227,560
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: 60/225,033
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 4
; LENGTH: 49744
; TYPE: DNA
; ORGANISM: Human
US-09-927-091-4

Query Match 91.1%; Score 16.4; DB 7; Length 49744;
Best Local Similarity 94.4%; Pred. No. 4.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 tgcagcagcgagcaagt 18
||||| |||||||
Db 35016 TGCAGCGCGCGCAGACT 34999

RESULT 12
US-09-724-671-4868/c
; Sequence 4868, Application US/09724671
; GENERAL INFORMATION:
; APPLICANT: Watson, James D
; APPLICANT: Murison, James G
; TITLE OF INVENTION: Polynucleotides, polypeptides expressed
; by the polynucleotides and methods for their use.
; FILE REFERENCE: 1050U2
; CURRENT APPLICATION NUMBER: US/09/724,671
; CURRENT FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: U.S. No. 60/171,678
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 21907
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4868
; LENGTH: 235
; TYPE: DNA
; ORGANISM: Mouse
US-09-724-671-4868

Query Match 88.9%; Score 16; DB 7; Length 235;

Best Local Similarity 100.0%; Pred. No. 5.5e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 gcagcagcgcgagaag 17
|||||
Db 96 GCAGCAGCGCGCAGAAG 81

RESULT 13

US-09-724-671-18201/c
; Sequence 18201, Application US/09724671
; GENERAL INFORMATION:
; APPLICANT: Watson, James D
; APPLICANT: Murison, James G
; TITLE OF INVENTION: polynucleotides, polypeptides expressed
; TITLE OF INVENTION: by the polynucleotides and methods for their use.
; FILE REFERENCE: 105002
; CURRENT APPLICATION NUMBER: US/09/724,671
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: U.S. No. 60/171,678
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 21907
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18201
; LENGTH: 326
; TYPE: DNA
; ORGANISM: Mouse
US-09-724-671-18201

Query Match 88.9%; Score 16; DB 7; Length 326;
Best Local Similarity 100.0%; Pred. No. 5.6e+02;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 gcagcagcgcgagaag 17
|||||
Db 89 GCAGCAGCGCGCAGAAG 74

RESULT 14

US-09-724-671-18461/c
; Sequence 18461, Application US/09724671
; GENERAL INFORMATION:
; APPLICANT: Watson, James D
; APPLICANT: Murison, James G
; TITLE OF INVENTION: polynucleotides, polypeptides expressed
; TITLE OF INVENTION: by the polynucleotides and methods for their use.
; FILE REFERENCE: 105002
; CURRENT APPLICATION NUMBER: US/09/724,671
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: U.S. No. 60/171,678
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 21907
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18461
; LENGTH: 338
; TYPE: DNA
; ORGANISM: Mouse
US-09-724-671-18461

Query Match 88.9%; Score 16; DB 7; Length 338;
Best Local Similarity 100.0%; Pred. No. 5.6e+02;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 gcagcagcgcgagaag 17
|||||
Db 89 GCAGCAGCGCGCAGAAG 74

RESULT 15

US-09-724-671-18790/c
; Sequence 18790, Application US/09724671

; GENERAL INFORMATION:
; APPLICANT: Watson, James D
; APPLICANT: Murison, James G
; TITLE OF INVENTION: polynucleotides, polypeptides expressed
; TITLE OF INVENTION: by the polynucleotides and methods for their use.
; FILE REFERENCE: 105002
; CURRENT APPLICATION NUMBER: US/09/724,671
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: U.S. No. 60/171,678
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 21907
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18790
; LENGTH: 363
; TYPE: DNA
; ORGANISM: Mouse
US-09-724-671-18790

Query Match 88.9%; Score 16; DB 7; Length 363;
Best Local Similarity 100.0%; Pred. No. 5.6e+02;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 gcagcagcgcgagaag 17
|||||
Db 89 GCAGCAGCGCGCAGAAG 74

Search completed: September 21, 2001, 02:39:19
Job time: 15045 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 21, 2001, 00:27:43 : Search time 3479.37 Seconds
(without alignments)
4338.867 Million cell updates/sec

Title: US-09-138-735-1_COPY_1232_2207
Perfect score: 976
Sequence: 1 cagtgacgcgcgtacgcgtt.....tgtgacgcgcgcattatgt 976

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba1: *
2: gb_ba2: *
3: gb_ba3: *
4: gb_in1: *
5: gb_in2: *
6: gb_in3: *
7: gb_om: *
8: gb_ov: *
9: gb_pat1: *
10: gb_pat2: *
11: gb_ph: *
12: gb_pl1: *
13: gb_pl2: *
14: gb_pl3: *
15: gb_pl4: *
16: em_ba1: *
17: em_ba2: *
18: em_fun: *
19: em_htgo_hum: *
20: em_htgo_inv: *
21: em_htgo_rod: *
22: em_htg_hum1: *
23: em_htg_hum2: *
24: em_htg_hum3: *
25: em_htg_hum4: *
26: em_htg_hum5: *
27: em_htg_hum6: *
28: em_htg_hum7: *
29: em_htg_hum8: *
30: em_htg_inv1: *
31: em_htg_inv2: *
32: em_htg_other: *
33: em_htg_rod: *
34: em_hum1: *
35: em_hum2: *
36: em_hum3: *
37: em_hum4: *
38: em_hum5: *
39: em_hum6: *
40: em_hum7: *
41: em_in: *
42: em_om: *
43: em_or: *

44: em_ov: *
45: em_pat: *
46: em_ph: *
47: em_pl: *
48: em_ro: *
49: em_sts: *
50: em_sy: *
51: em_un: *
52: em_vt: *
53: gb_sts1: *
54: gb_sts2: *
55: gb_sts3: *
56: gb_sy: *
57: gb_un: *
58: gb_vt1: *
59: gb_vt2: *
60: gb_vt3: *
61: gb_vt4: *
62: gb_vt5: *
63: gb_vt6: *
64: gb_vt7: *
65: gb_vt8: *
66: gb_vt9: *
67: gb_vt10: *
68: gb_vt11: *
69: gb_vt12: *
70: gb_vt13: *
71: gb_vt14: *
72: gb_vt15: *
73: gb_vt16: *
74: gb_vt17: *
75: gb_vt18: *
76: gb_vt19: *
77: gb_vt20: *
78: gb_vt21: *
79: gb_vt22: *
80: gb_vt23: *
81: gb_vt24: *
82: gb_vt25: *
83: gb_vt26: *
84: gb_vt27: *
85: gb_vt28: *
86: gb_vt29: *
87: gb_vt30: *
88: gb_vt31: *
89: gb_vt32: *
90: gb_vt33: *
91: gb_vt34: *
92: gb_vt35: *
93: gb_vt36: *
94: gb_vt37: *
95: gb_vt38: *
96: gb_vt39: *
97: gb_vt40: *
98: em_ba3: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	976	100.0	3402 9	AR047920
2	972.8	99.7	3402 9	A48910 Sequence 1
3	972.8	99.7	3402 96	TC024190
4	59.6	6.1	40309 6	LMF11923
5	59.6	6.1	110000 84	LMF1CHR31_04
6	47.8	4.9	2402 3	MLURHO
7	46.6	4.8	175105 61	AC010531
8	46.6	4.8	179034 61	AC009109 Homo sapi

Accession	Gene	Species	Length (bp)
AL020222	Myocobacter	Myocobacter	1355
AL021841	Myocobacter	Myocobacter	1355
AJ132828	Spermatoz	Myocobacter	1355
AC008532	Homo sapi	Myocobacter	1355
M30933	E.tenella	Myocobacter	1355
AP001081	Oryza sat	Myocobacter	1355
D14486	Equine herp	Myocobacter	1355
U87960	Rattus norv	Myocobacter	1355
AC008735	Homo sapi	Myocobacter	1355
E54519	scat gene	Myocobacter	1355
AF020275	Homo sapi	Myocobacter	1355
AI133030	Homo sapi	Myocobacter	1355
X83546	R.norvegicu	Myocobacter	1355
AB051453	Homo sapi	Myocobacter	1355
Z81368	Myobacteri	Myocobacter	1355
AL009138	Myocobacter	Myocobacter	1355
AP000556	Homo sapi	Myocobacter	1355
AP000557	Homo sapi	Myocobacter	1355
AP000552	Homo sapi	Myocobacter	1355
AC018751	Homo sapi	Myocobacter	1355
AC009516	Homo sapi	Myocobacter	1355
AC058816	Homo sapi	Myocobacter	1355
AC023490	Homo sapi	Myocobacter	1355
AC007957	Homo sapi	Myocobacter	1355
AF310215	Sorghum b	Myocobacter	1355
AL021897	Myocobacter	Myocobacter	1355
AP003297	Oryza sat	Myocobacter	1355
AC044842	Homo sapi	Myocobacter	1355
AF139019	Cepaea ne	Myocobacter	1355
E532986	Gene encod	Myocobacter	1355
AC090433	Chlamydom	Myocobacter	1355
AC090650	Arabidops	Myocobacter	1355
X17207	Chlamydomon	Myocobacter	1355
AI143840	T3 end of	Myocobacter	1355
AF087653	Homo sapi	Myocobacter	1355
AL446003	Streptomy	Myocobacter	1355
U138713	Gallus gall	Myocobacter	1355

Qy	661	gcacacgataaccagctcgagagggcgaggttcatcaatgcctcgaagcgctctgaagtccttc	720
Db	1892	gcacagctatnaccagtcggagggcgaggttcantcagctcgtcgagcgctctgaagtccttc	1951
Qy	721	ataaccacaagctctcggtcttgaagtgttcctctgaaccattcgaagagacatgaa	780
Db	1952	ataaccacaacgctcgcggtttcgaagttgctcccgacacatttcgaagacacatgaa	2011
Qy	781	caactcttgatctcgggttttagagagcaatagaagagctcgaagagagcgcttcaaca	840
Db	2012	caactcttgatctcgggttttagagagcacaagatcacagagctgcagcagagccgtccmaca	2071
Qy	841	ccgacaacaagccgagagacacaagaagctccgagaaatcatccgtgtgttgaagcgtacac	900
Db	2072	ccgacaacaagccgagagacacaagaagctccgcaaatcattccgtgtttgaagcagtaaac	2131
Qy	901	ctgttctcattgagatctccctctcgcgaacaatacgaagaggggtgaagctgtgtgt	960
Db	2132	ctgttctcattgagagattccctctctcgcacatacacgaaggggtgaagcgtgtgtgt	2191
Qy	961	aacgagacatlatgt 976	
Db	2192	aacgagacatlatgt 2207	

[illegible]

REFERENCE	Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma; Schizotrypanum.
AUTHORS	1 (bases 1 to 3402)
TITLE	Paranhos-Baccala,G., Lesenechal,M. and Jolivet,M. NOVEL TRYPANOSOMA CRUZI ANTIGEN, AND GENE CODING APPLICATION TO THE DETECTION OF CHAGAS' DISEASE
JOURNAL	PATENT: WO 9605312-A 1 22-FEB-1996; BIO MERIEUX (FR)
COMMENT	Other publication CA 2173957 960222 Other publication AU 3169195 960307 Other publication FR 2723589 960216.
FEATURES	Location/Qualifiers
SOURCE	1. 3402
	/organism="Trypanosoma cruzi"
	/strain="G"
	/db_xref="taxon:5693"
	/dev_stage="EPIMASTIGOTE"
BASE COUNT	889 a 818 c 958 g 737 t
ORIGIN	

Query Match	Similarity	99.7%	Score	97.8	DB	9	Length	3402
Best Local	Similarity	99.8%	Pred	No.	7.8e-203			
Matches	974	Conservative	0	Mismatches	2	Indels	0	Gaps
QY 1	caggtacagcgttaacgagcttctgtctcaactcgttaacacgcagcgttagctctgctgact	60						
Db 1232	CAGGTACAGCGTAAACGGCGTTTGTCTTCAATCGTACAGCCGACGGATGACGTGCGTGGCT	1291						
QY 61	gacatgtcgtatctgatcttgaagatctcttcaactctccgtgtctcccgcaaggaaacagcag	120						
Db 1292	GACATGTGTAATCGATTACGATTCATTCATCCGCTCTCCCGACGGATMAACAGCGAG	1351						
QY 121	ccaggtcccaaaaacacatcggttagtggcagcagggaaacccgggtgtgtgtccctcgagcact	180						
Db 1352	CCAGGCCCAAAAAACATCTCGTACTGGCGACAGCGGAAACCCGGGTGTGTCTCTCGGCACT	1411						
QY 181	gacgcgcgcagtagtagcagtcataccaatacgaactctgtccgcgtgtgtgcatccctctgataca	240						
Db 1412	GACGCGGCGAGTAGCAGTCACTATCCAAATACGACTTCTGCGCGTGTGATCCCTGCATCA	1471						
QY 241	ccccctgttttaagcgcgcagcccaaggcagcgcgcgtctctgtccggtgcgcgtacgtgtgag	300						
Db 1472	CCCCCTGTTTACGCGCCACGCGAAGGACGCGCGCTCCGCGCGCGCGGATTCGGCTGAG	1531						
QY 301	ccgcacgtgtgggagaaagatactgtctaactcagtgaaatcagctgtgggattaatgtacc	360						
Db 1532	CCGCACGTGGGGAGCAAGTCACTTGTCTAATCTAGTGAATCACTGGGATTAATGTACCC	1591						
QY 361	caaaagagcgtctgtcagcacttgaagcgcgcgcgcacagcaggttctacgcggtgtgactgcc	420						
Db 1592	CAAAAGAGACGTGTGTGACGACTGTGAGCGCGCGCCACGACGAGGTCTACGGCGGTGACGTCC	1651						
QY 421	acgactaccgcgcgcgcagcagcgaacagtctcatactcagggcacaaatgtgcgcactgttagcgtc	480						
Db 1652	ACGACTACCGCGCCCGCCACGAAACAAGTCCATACGGGCACAATAGCGCGACTGTGAGCGCT	1711						
QY 481	ggattgtgtgcgcacatctagtgtgtgcagcgcgcgcgtctgtccccaacacgcgcgggaaa	540						
Db 1712	GGATTGTGTGCGACTTAATATGTGTGTCACGCGGGCGCTCTCTCCCAACACCGCGGGGAAA	1771						
QY 541	ccaacaagagagaanaaagccttcgcgcgcgcgtgtgaacacagcttccgtgtgcgataaattgcg	600						
Db 1772	CCAACAAGAGGAAGAAAAGGCGCTCCCGCGGATGTGAACAGCAGCTCGGTGGCGATTAATGGC	1831						
QY 601	acaagcccgcggtcttcaacaagcctctctcccgagcgcgcaacggaatgtgcgttttggtcg	660						
Db 1832	ACACCCCGCGGCGCTTCAACAACCTCTCTCCCGCAGGCGCCAAACGATGGCGTTTGGCGC	1891						
QY 661	ggaagagatatacgaatctggaggagcaggttatacagttcgtctgtgagggtgtgagtgcgtc	720						
Db 1892	GGAGAGATATACCAAGTCGAGGGCGACAGGTTCATCATCTCTGAGCGCGCTGAGATCCGTC	1951						

QY	721	ataccaaacagctcgcgggtcttctgaagttgcctcccgaaacattctgaagaagacatgaa	780
Db	1952	ATAACCAACAGCTCGCGGTTCTGAAATTCTCTCCGTGACCACTTCGGAAGAACCATGTAA	2011
QY	781	caactctgaa tctgggtctttagagcagatgacagagctgcagcagagccgtccaa	840
Db	2012	CAACTCTGAA TCTGGGTTTAGAGCAGACATGACAGAGCTGCAGCAGACCGCTCAACA	2071
QY	841	ccgcgaacaacgcccgaagagacaacagctccgcggaatcatcgttgtttagacgtacac	900
Db	2072	CCGCAAAACACGCCGAGAAGACACAAAGCTCCGCCAATAATCATCCGTGTTTAAAGACGTAACAC	2131
QY	901	cttgtcttcattgcyggaattccacctctcgcgaacatcaagaaaggggttgaagcgtgtgtg	960
Db	2132	CTTGTTCTCATTTGGGGAATCCCTCTCTCGCAACATCACGAAGGGGCTGAAGCTGTGTG	2191
QY	961	aacgagagccattatgt 976	
Db	2192	AACGAGGCCATTATGT 2207	

RESULT	3
LOCUS	TCU24190
DEFINITION	Trypanosoma cruzi Tc40 bp mRNA (Tc40) mRNA, complete cds.
ACCESSION	U24190
VERSION	U24190.1 GI:790645
KEYWORDS	
SOURCE	Trypanosoma cruzi.
ORGANISM	Trypanosoma cruzi Eukaryota; Euzoenozoa; Kinetoplastida; Trypanosomatidae;
REFERENCE	Trypanosoma; Schizotrypanum. 1 (bases 1 to 3402) Lesenechal,M., Duret,L., Cano,M.I., Mortara,R.A., Jolivet,M., Camargo,M.E., da Silveira,J.F. and Paranhos-Baccala,G. Cloning and characterization of a gene encoding a novel immunodominant antigen of Trypanosoma cruzi Mol. Biochem. Parasitol. 87 (2), 193-204 (1997)
JOURNAL	2 (bases 1 to 3402) Lesenechal,M., Franco Da Silveira,J., Mortara,R.A., Duret,L., Camargo,M.E., Jolivet,M. and Paranhos-Baccala,G. Direct Submission Submitted (06-APR-1995) Mylene Lesenechal, Biomerieux, 22 rue Saint Jean de Dieu, Lyon 69007, France
MEDLINE	Location/Qualifiers 1..3402
AUTHORS	/organism="Trypanosoma cruzi" /strain="G" /db_xref="taxon:5693" /dev_stage="epimastigote" 1..35 /gene="Tc40" /note="spliced leader" 1..3402 /gene="Tc40" 266..3013 /gene="Tc40" /codon_start=1 /product="Tc40 antigen" /protein_id="AAC47657.1" /db_xref="GI:790645"
FEATURES	translation="MTATVDFLFNNAKPSNNEGRWASVDAAFTFNEPEAORVLADSOFY LAYMKRRHVRVVYKRSNLKGTVBARSKPLHAVKFVNSRNAASAGKEFFWAVYT DEFASSMGKDLARILTYKKVEFKLDDPVTTIPCFSEFFINAESORPDLIVLETOAATID SSSLIERDVESLEATLRQNTTIRLTLPQVSENSLCVSQGGFTTTPPTVAACT LRNSTBSMACCSGEPEPAVKALHLDAFLAEENVSVLVASTGVQWLTVGAEPNLARK FVIDGSIVAMESRETEPAVEDGRDKRLDALVNHHSHNTCTHYMPCCOVONGCFFNT ADGSCVLADMSNRLLTIHLKSREEOOPCKTSVTATARPQCVASSGTDASSHNTY TSAASAASPVPSPAKAAPAAAPAAASAEPHSGKITALLVNOGLINTVORSVNG APATTBRNAVSTTTTAQRQRTSPYGHNRPRPTAGLVANSQSASASPTAAKPTGEK ASAKCEITSSVAINTREPALHNASLPQAPTDGVLAAAYOSGEVHOSLEAFLESVITWT SRVALCPDTTRRHQDLMLNGLEAMQETLEOOSPRTPOTPRPDTSSAKSSSEFTYTLV
misc_RNA	
gene	
CDS	

LADSLSRNITGKVRGNBEALMLHIDHEVRHAIKRLROTKNITKSRIDEALKEST
 TOFTAOITQVENVLRKLAELVGLSINGSLTSLVKCNALQKLNINSVGVDEMR
 MREELCTRESVAKRKRAIPDSSLIATSSFOGRSRSPETILATLSMSEBOQROGLE
 YMLMAQPSLLREFLSILTRFNENAYSELLENWCVSLLOLIDLEAAAEAKE
 VYVGAVIDLSEBDQIAONGALSGSKLTITMRAFEERARSETTSRSRLQCKLNKEKLLO
 S"

BASE COUNT 889 a 818 c 958 g 737 t
 ORIGIN

Query Match 99.7%; Score 972.8; DB 96; Length 3402;
 Best Local Similarity 99.8%; Pred. No. 7.8e-203;
 Matches 974; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 1 caggtacaggtacagcgttttgtctcaatcgtacagcagcaggtacgtcgtctgct 60
 |||||
 Db 1232 CAGGTACAGGCTAACGGCTTTTGCTTCAATCGTACACCGACGGTACTCGTCTGGCT 1291
 |||||
 Db 61 gacatgtcagatcgtatgacgattctccatctccggtctcccgaggaagacagcag 120
 |||||
 Db 1292 GACATGTGCAATGCAATGACGATCTTCATCTCCGCTCCGCCAGGAAGAACACAG 1351
 |||||
 Db 121 ccagggccaaaacatcgtcgtatgtggcagcagcagcagcagcagcagcagcagc 180
 |||||
 Db 1352 CCAAGCCAAAACATCGGTAGTGGCGACGCGAACCAGGAGGTGTGTCTCGGGCACT 1411
 |||||
 Db 181 gacgagcagatagcagtcataacacacacacacacacacacacacacacacacac 240
 |||||
 Db 1412 GAGCGGCGAGTACAGTACATACCAATACACATCTTCCGCTGTGATCCCGTCATCA 1471
 |||||
 Db 241 ccccttgcttcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 300
 |||||
 Db 1472 CCCCCTGTTTCACGCGCCAGGAGGACGCGCGCTCTCCGCGCGCGCATCGCGTAG 1531
 |||||
 Db 301 ccgacagctgggagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 360
 |||||
 Db 1532 CCGCAGCTGGGGACCAAGATCTTGTCTAATCTAGTGAATACGCTGGGGATTAAATGTCACC 1591
 |||||
 Db 361 caaagagcgtcgtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 420
 |||||
 Db 1592 CAAAGGAGCGTGTACACACTGTGAGCGCGCGCCACGAGAGGTCTACGGCGTACGCTCC 1651
 |||||
 Db 421 agcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 480
 |||||
 Db 1652 ACAGACTACCCCGCGCGCAACAAGTCATACGCGGACATGCGCCACCTGTGACGGCT 1711
 |||||
 Db 481 ggaattgtgagcagtaatagtgtgagcagcagcagcagcagcagcagcagcagcagcag 540
 |||||
 Db 1712 GGATTGTGGCAGCTAATATGTGTGCGACGCGCGCTCTCTCCACAGCGCGCGCAAA 1771
 |||||
 Db 541 ccaacaggaagaaagagcctccgcgagcagcagcagcagcagcagcagcagcagcagcag 600
 |||||
 Db 1772 CCAACAGGAGAAAGAGCGCTCCGCGGAGTGTGAAGAGAGCTCCGTGGCGATTAATGCGG 1831
 |||||
 Db 601 aacagcccgagcgtctcaacagcgtctctccgagcagcagcagcagcagcagcagcagc 660
 |||||
 Db 1832 ACAGCGCCCGGCTTCAACAGCGCTCTCTCCGAGCGCAACGAGTGGCGTTTGGCG 1891
 |||||
 Db 661 ggcagcagatacagcgtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 720
 |||||
 Db 1892 GCAGCAGTATACAGTGTGAGGCGGAGGAGTTCATCAGTGTGAGGCGGCTGTGAGTCCGCTC 1951
 |||||
 Db 721 ataaccaaacagctcgtggtcttgaagtgtcccttgcacacacacacacacacacacacac 780
 |||||
 Db 1952 ATTAACCAACAGCTCTCGGCTTGTGAAGTGTCTCCCTGACACACATTCGAAGAGACATGAA 2011
 |||||
 Db 781 caacttcgatcttggttttagagagcagatgacagagctgacagcagcagcagcagcagcagcag 840
 |||||
 Db 2012 CAACCTTGTGATCTGGGTTTGAAGGCAACATGACAGAGTGCAGCAGACCGCTGCCAACA 2071
 |||||
 Db 841 ccgcaaacacagccgagagacacagctccggaatcacctcagcagcagcagcagcagcagcagcagc 900
 |||||
 Db 2072 CCGCAAAACACAGCGAGACACAGACTCCGCGAAATCATCTCCGTGTTGAGAGCTTACACC 2131

QY 901 ctgtctcattcgcggaattccctctctcgcacacacacagagggggtgaagcgttggtg 960
 |||||
 Db 2132 CTGTCTCTAATGTTGGGATTCCTCTCTCCGACATACAGGAAGGGGTGAAGCGGTGTG 2191
 |||||
 QY 961 aacgagccattatgt 976
 |||||
 Db 2192 AACGAGCGCATATGCT 2207

RESULT 4

LMF1923 40309 bp DNA INV 16-MAY-2000
 LOCUS Leishmania major Friedlin chromosome 14 cosmid L1923.
 DEFINITION AL356246
 VERSION AL356246.1 GI:7940262
 KEYWORDS possible DNA polymerase I-related protein; possible glutathione synthetase; possible ribosomal protein; possible rC40 antigen; protein phosphatase 2C-like protein.
 SOURCE Leishmania major
 ORGANISM Eukaryota; Eukaryota; Kinetoplastida; Trypanosomatidae; Leishmania.
 REFERENCE 1 (bases 1 to 40309)
 AUTHORS Ivens,A.C., Lewis,S.M., Bagherzadeh,A., Zhang,L., Chan,H.M. and Smith,D.F.
 TITLE A physical map of the Leishmania major Friedlin genome
 JOURNAL Genome Res. 8 (2), 135-145 (1998)
 MEDLINE 98146435
 REFERENCE 2 (bases 1 to 40309)
 AUTHORS Murphy,L., Harris,D., Ivens,A.C., Quail,M., Rajandream,M.A. and Barrell,B.G.
 TITLE Direct SubMISSION
 JOURNAL Submitted (16-MAY-2000) European Leishmania major Friedlin genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, (E-mail: barrell@sanger.ac.uk)
 COMMENT see <http://www.ebi.ac.uk/parasites/leish.html>
 DETAILS: Details of Leishmania sequencing at the Sanger Centre are available on the World Wide Web.
 see <http://www.sanger.ac.uk/projects/Lmajor/>
 CDS are numbered using the following system eg L1923.01. L1923 (cosmid name), .01 (first CDS)
 To make the cosmid library Leishmania major Friedlin DNA was partially digested with Sau3AI prior to cloning into BamHI site of the cosmid shuttle vector cLHYG (Ryan et al. 1993 Gene 131:145-150). The sequence of the packaged vector was determined by Peter Myler and Ken Stuart at Seattle Biomedical Research Institute, and is available as accession number U59231.
 The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.
 The length in codons is given for each CDS.
 Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database.

Gene prediction is done using:

(1) the Frameplot program of Bibb et al.,
 Gene 30:157-166(1984) as implemented
 at <http://www.nih.go.jp/jun/cgi-bin/frameplot.pl>. (2)
 codon preference based on the codon usage table for Leishmania at
<http://www.kazusa.or.jp/codon/>
 (3)

the Hexmer program which was written by Richard Durbin as an integral part of the ACEDB-based analysis tools for the C.elegans Genome Sequencing Project. The program calculates the frequency log-likelihood score for a given DNA segment based on the frequency of 6-mers, normalised for the base-pair composition of the genome. The program was trained using a fasta file of confirmed Leishmania major coding sequences (CDS), i.e. from ATG start codon to the stop codon.
 CAUTION: We may not have predicted the correct initiation codon.

Db 12568 CCTGTGCGCTTCGCTCCGCCGCAACGACGCTGCTCGACCGCTTCGAACGGCAA 12627
QY 570 atgtgaacgagctcgttggcgaataatgacagccgagcgttcacacgctctct 629
Db 12628 CCGGATTAACGATGCGCGGCGGCAACAGCGGCTCGGACGGCTCTCAAAACGGCACCT 12687
QY 630 cccgacgagcgcaacgagatgcttctgagcgcagcagatcaccagtcgagagcgagt 689
Db 12688 GGCACACATGCTCC---TGGACGCGGCTTGTGGAGCGCGCTTACACAGGCGACGGCCCTT 12744
QY 690 tcatcagctgctgagagcggtgtgagtcctcacaacacagctctcgtggtctgaagt 749
Db 12745 GCGCGCGCGCATGATGATGCTGCTGCTGATGTTTGTAAGAACGTTACGAGATCTGTTGCT 12804
QY 750 gctccctgacacattcgaagagacacatgaacacatctgtaattcgtggtttagagaca 809
Db 12805 CACCCCGATGACGCTCAGCGCGGATCAGCAGCAGCTGCTGCTCACTGCGGTGGAGGGCA 12864
QY 810 gatgacagagctgcagcagagc 831
Db 12865 GATGACGAGACTGACAGCAGC 12886

RESULT 5

LMFLCHR31_04
WPCOMMENT

Sequence split into 11 fragments LOCUS LMFLCHR31 Accession AL499621

Fragment Name	Begin	End
LMFLCHR31_00	1	110000
LMFLCHR31_01	100001	210000
LMFLCHR31_02	200001	310000
LMFLCHR31_03	300001	410000
LMFLCHR31_04	400001	510000
LMFLCHR31_05	500001	610000
LMFLCHR31_06	600001	710000
LMFLCHR31_07	700001	810000
LMFLCHR31_08	800001	910000
LMFLCHR31_09	900001	1010000
LMFLCHR31_10	1000001	1027083

Continuation (5 of 11) of LMFLCHR31 from base 400001 (AL499621) Leishmania major chromosome

Query Match 6.1%; Score 59.6; DB 84; Length 110000;
Best Local Similarity 51.2%; Pred. No. 0.0012;
Matches 165; Conservative 0; Mismatches 154; Indels 3; Gaps 1;

QY 510 cgcgagcctgctcccaacgacgcgcgcgaacacacagagagaagaagcctccgcgcgc 569
Db 63000 CCTGTGCGCTTCGCTCCGCCGCAACGACGCTGCTCGACCGCTTCGAACGGCAA 63059
QY 570 atgtgaacgagctcgttggcgaataatgacagccgagcgttcacacgctctct 629
Db 63060 CCGGATTAACGATGCGCGGCGGCAACAGCGGCTCGGACGGCTCTCAAAACGGCACCT 63119
QY 630 cccgacgagcgcaacgagatgcttctgagcgcagcagatcaccagtcgagagcgagt 689
Db 63120 GGCACACATGCTCC---TGGACGCGGCTTGTGGAGCGCGCTTACACAGGCGACGGCCCTT 63176
QY 690 tcatcagctgctgagagcggtgtgagtcctcacaacacagctctcgtggtctgaagt 749
Db 63177 GCGCGCGCGCATGATGATGCTGCTGCTGATGTTTGTAAGAACGTTACGAGATCTGTTGCT 63236
QY 750 gctccctgacacattcgaagagacacatgaacacatctgtaattcgtggtttagagaca 809
Db 63237 CACCCCGATGACGCTCAGCGCGGATCAGCAGCAGCTGCTGCTCACTGCGGTGGAGGGCA 63296
QY 810 gatgacagagctgcagcagagc 831
Db 63297 GATGACGAGACTGACAGCAGC 63318

RESULT 6

MLJRH0

LOCUS MLJRH0 2402 bp DNA BCT 12-NOV-1996
DEFINITION Micrococcus luteus Rho factor (rho) gene, complete cds.
ACCESSION L27277
VERSION 1.27277.1 GI:1666539
KEYWORDS
SOURCE Micrococcus luteus.
ORGANISM Micrococcus luteus
Bacteria: Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Micrococcales; Micrococcaceae; Micrococcus.
1 (bases 905 to 2402)
REFERENCE
AUTHORS Oppertman,T. and Richardson,J.P.
TITLE Phylogenetic analysis of sequences from diverse bacteria with
homology to the Escherichia coli rho gene
J. Bacteriol. 176 (16), 5033-5043 (1994)
94327472
JOURNAL 2 (bases 1 to 2402)
MEDLINE Nowatzke,W.L. and Richardson,J.P.
REFERENCE Characterization of an unusual Rho factor from the high G + C
J. Biol. Chem. 271 (2), 742-747 (1996)
96132802
AUTHORS Richardson,J.P.
TITLE Direct Submission
JOURNAL Submitted (14-FEB-1994) John P. Richardson, Chemistry, Indiana
REFERENCE University, Bloomington, IN 47405, USA
AUTHORS Nowatzke,W.L.
TITLE Direct Submission
JOURNAL Submitted (07-NOV-1996) William L. Nowatzke, Chemistry, Indiana
REFERENCE University, Bloomington, IN 47405, USA
COMMENT On Nov 12, 1996 this sequence version replaced gi:968908.
FEATURES
source
1..2402
/organism="Micrococcus luteus"
/strain="EM"
/db_xref="taxon:1270"
288..2363
gene
288..2363
CDS
288..2363
/gene="rho"
/function="transcription termination factor"
/codon_start=1
/transl_table=11
/product="Rho factor"
/protein_id="AAB18671.1"
/db_xref="GI:1666540"
/translation="MRESEORTPTNGGSLAKIQAOLALASOLGIAGSRMRADL
VRAISHQSGSVADRDADAEARAQAAPAAATAPAAASSEDAAPAEAPARRRSRA
DADTSAFPAADQDQPOAEAREKOTEDAPRETAISDDRSSEARDEGEDEPSEERSR
GRRACDDDAQDQDSDGACQEDGADADRGRDRDRDNGRENGRGRNGRGRDRD
NGRDRNGRENGRNGRNGRNGRNGRNGRNGRNGRNGRNGRNGRNGRNGRNGRNGR
NRNRNRNGRNGRNGRNGRNGRNGRNGRNGRNGRNGRNGRNGRNGRNGRNGRNGR
VYVSLAMVKYGRKGDVAVGPAPADGKQOHGGSGNRKFNALVKTSSVNGOPAV
EHPOREREKLPVLPORERLLETQPKLIGPRVIDVSPIGGONRLVISPCKAKTAM
IIQSTINAKTKNNPEVHLMLVLYDERPEEVTMORSYDEVIASFTDRADHTTLAE
LAIERAKRLVEKGRDVLVLDSTWTRGLRVATLAAPRSALISGVDSSALVPPKFTG
LAARNIENGSLTTLATVETGSRMDEVFEFEKGGNMLRLSRHLAERLFPADV
NASGTREREAALLSDEEVKIMWKLRLVLSGLEDDQDALDLTKNKIKDYASNAEFLMLVSK
TTLGSKGDD"

BASE COUNT 386 a 947 c 818 g 251 t
ORIGIN

Query Match 4.9%; Score 47.8; DB 3; Length 2402;
Best Local Similarity 50.2%; Pred. No. 1.5;
Matches 118; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

QY 367 agcgtcgtcagcactgagcgccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 426
Db 799 AGCAGGGTCAAGACCGTCTCGACGCGCGCCGAGGCGAGGCGCGCGCGCGGACC 858
QY 427 accgccccgcagcaacaagtcacacgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 486

Db 859 GCGCGGACCGGAGACCGTGACGACAGCGGCGCGAGACGCGCGCGGCAACG 918
Qy 487 gtggcagctaatatgtgtgcagcgcgccctcgtctcccaagcgcggaacaaca 546
Db 919 GCCCAACGGCGGACCGGACGACGCGCGCGGACGAGAACGCGCGGCAACGCC 978
Qy 547 ggaagaagaagcgctccgcgcatgtgaacgagctccgtgagataatgcga 601
Db 979 GGGACCGTGAGAACGCGCGGACGCGCTCGCGGAGCAGCGGCGGACAGTCCGA 1033

RESULT 7
AC010531
LOCUS Homo sapiens chromosome 16 clone RP11-178L8, WORKING DRAFT
DEFINITION AC010531
ACCESSION AC010531.5 GI:9954623
VERSION HTGS_PHASE2; HTGS_DRAFT.
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 175105)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 16
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 175105)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Aug 31, 2000 this sequence version replaced gi:8576022.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
Genbank flat file format but are available as part
of this entry's ASN.1 file.

* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 151748: contig of 151748 bp in length
* 151749 151848: gap of unknown length
* 151849 175105: contig of 23257 bp in length.
Location/Qualifiers
1. 175105
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-178L8"
/clone="RP11-178L8"

BASE COUNT 47958 a 43943 c 43102 g 40002 t 100 others
ORIGIN

Query Match 4.88; Score 46.6; DB 61; Length 175105;
Best Local Similarity 49.4%; Pred. No. 0.76;

Matches 121; Conservative 0; Mismatches 124; Indels 0; Gaps 0;
Qy 538 aaaccaagaagaagaagcctccgcgcatgtgaacgagctccgtgagataat 597
Db 125630 AAAACAAAACAAAACAAATGGTCAACGTCACCAAAACCAACAGAGTCAACAACT 125689
Qy 598 gcgaacgcccgcgcttcaacacgctctctccgcgagcgcaagatgaggtttg 657
Db 125690 GTCACAGCCCGAGAGAGGACACGAGGACTGAATGCCAGGGGGCTCCAGATGGGATTCG 125749
Qy 658 gcgcagcagatataccagtcgagagcgaggttcacagtcgctgagcgagtc 717
Db 125750 GGAGAGAAAAGATGTGATGAAACACGCGGAATCTGAACCAAGTGCAGCGTGTTA 125809
Qy 718 gtcataaccacacgctcgggttcgaagttgctccctgacaccatcgagaagacc 777
Db 125810 ATCATACACATCATATTATTGTTGTTAACTGACCAATATGCAATCTAAGAGAAC 125869
Qy 778 gaaca 782
Db 125870 GGTCA 125874

RESULT 8
AC009109/c
LOCUS Homo sapiens chromosome 16 clone RP11-467L24, WORKING DRAFT
DEFINITION AC009109
ACCESSION AC009109.8 GI:9954606
VERSION HTGS_PHASE2; HTGS_DRAFT.
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 179034)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 16
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 179034)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Aug 31, 2000 this sequence version replaced gi:9256136.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
Genbank flat file format but are available as part
of this entry's ASN.1 file.

* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 23258: contig of 23258 bp in length
* 23259 23358: gap of unknown length
* 23359 31091: contig of 7733 bp in length
* 31092 31191: gap of unknown length

```

* 31192 66150: contig of 34959 bp in length
* 66151 66250: gap of unknown length
* 71243 71243: contig of 4993 bp in length
* 71244 71343: gap of unknown length
* 71344 84814: contig of 13471 bp in length
* 84815 84914: gap of unknown length
* 84915 100309: contig of 15395 bp in length
* 100310 100409: gap of unknown length
* 100410 107673: contig of 7264 bp in length
* 107674 107773: gap of unknown length
* 107774 109247: contig of 1474 bp in length
* 109248 109347: gap of unknown length
* 109348 123504: contig of 16257 bp in length
* 125605 125704: gap of unknown length
* 125705 144649: contig of 18945 bp in length
* 144650 144749: gap of unknown length
* 144750 174670: contig of 29921 bp in length
* 174671 174770: gap of unknown length
* 174771 176776: contig of 1906 bp in length
* 176777 176776: gap of unknown length
* 176777 177942: contig of 1166 bp in length
* 177943 178042: gap of unknown length
* 178043 179034: contig of 992 bp in length.
* Location/Qualifiers
  1. 179034
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="16"
    /clone="RP11-467L24"
BASE COUNT 40736 a 43378 c 44428 g 49190 t 1302 others
ORIGIN

Query Match      4.8% Score 46.6; DB 61; Length 179034;
Best Local Similarity 49.4%; Pred No. 0.75;
Matches 121; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 538 aaacacacagagagaaagagcctccgcgcagcgtgaaacagagcgtcgltgcgataaat 597
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 48920 AAAACAAACAAAACAAAATCGCTCAGCTCAGCAACAAACAGAGAGCTCTAACAACT 48861

QY 598 ggcagacgcgcgcgcgtctacacagcctctctccgcagcgccagcgatcgctttg 657
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 48860 GTCACAGCCCAAGAGGAGCGACGACTGAATCCACGGGGCTCCCAATGGGATTCG 48801

QY 658 ggcgcagcagcgtacacagtcgcagcgaggtcaccagtcgcgtgcgcgcgcgcgc 717
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 48800 GGAGAGAAAGAGATGTGACTGAAACCTCGGAATCTGACCAACTGCACAGCGTGTG 48741

QY 718 gtcataacacacagtcgtcggttctgaagtgtctccctgcacacacattcgaaagacat 777
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 48740 ATCATTAACACATCATTAATGTTTGTACTGTGACCAATATGCTACTAAGAGAAAGAC 48681

QY 778 gaaca 782
    |||||
DB 48680 GGTCTA 48676

```

```

RESULT 9
MTV023      47852 bp      DNA      BCT      17-JUN-1998
DEFINITION  Mycobacterium tuberculosis H37Rv complete genome; segment 148/162.
ACCESSION  AL022022 AL123456
VERSION    AL022022.1 GI:3261554
KEYWORDS
SOURCE
  ORGANISM  Mycobacterium tuberculosis.
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacteriaceae; Mycobacteriaceae;
            Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE  1 (bases 1 to 47852)
            Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C.,
            Harris, D., Gordon, S.V., Eigemeier, K., Gas, S., Barry III, C.E.,

```

```

TITLE
JOURNAL
MEDLINE
REMARK
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Notes:
Details of M. tuberculosis sequencing at the Sanger Centre are
available on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/M.tuberculosis/) CDS have
been renumbered from the original cosmid submissions but the old
gene designations are in brackets after the new gene numbers.
Gene prediction was based on a Hidden Markov Model of TB genes
Implemented in tparse (Krogh) supplemented with visual inspection
of positional base preference in codons, especially where there is
an increase in the observed/expected third position G + C.
CAUTION: In some cases we may not have predicted the correct
initiation codon. Where possible we choose an initiation codon
(atg, gtg, or ttg) which is preceded by an upstream ribosome
binding site sequence (optimally 5-13bp before the initiation
codon). If this cannot be identified we choose the most upstream
initiation codon.
Location/Qualifiers
  1. 47852
    /organism="Mycobacterium tuberculosis"
    /strain="H37Rv"
    /db_xref="taxon:1773"
    <1..18
    /organism="Mycobacterium tuberculosis"
    /strain="H37Rv"
    /db_xref="taxon:1773"
    /clone="Y13E12"
    <1..47852
    /note="fragment designated V023. Does not represent a
    physical clone"
    complement(1..1695)
    /gene="RV3494c"
    complement(1..1695)
    /note="RV3494c" (MTV023.01c), len: 564. Unknown Pro-rich
    protein similar to several Mycobacterium tuberculosis
    proteins e.g. MTC128.14 (515 aa), MTCY19H5.28C (516 aa) and
    (MTV051.09). Has hydrophobic stretch, possibly signal
    peptide at N-terminus. Fasta scores: 2970501MTC128.14
    (515 aa) Opt: 1006 z-score: 582.8 E(1): 5.9e-25;
    identify in 559 aa overlap: 2971821MTCY19H5.28 (516 aa)
    Opt: 979 z-score: 567.7 E(1): 4.1e-24; 33.5% identity in
    555 aa overlap. Tparse score is 0.897".
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein RV3494c"
    /protein_id="CA117731.1"
    /db_xref="GI:2924431"
    /db_xref="SPTREMBL:O53539"
    /translation="MIDRLAKTOLSTIFAVITVITLSVAIFRLRLPATEFGIGYVSA
    DEVAGGGLKRNANVTTRGVANGRVSGVGNPGVTAHNRKLNSGRTAIPSNVATRAVS
    AIGEYIDLVPPENSTKLKNGRFRIORONTRIGODVADLLRQAEITLGSIGDRFLRE

```

```

FEATURES
  source
    misc_feature
    gene
    CDS

```


Oy		538	aaaccaacaggaagaaggcccccgcgqga	570
Dn	32922	ACGCCGCGCCGGAGCAATGCACAACGCGCAGA	32954	

RESULT 10

LOCUS	MWV016	53662 bp	DNA	BCT	23-JUN-1999
DEFINITION	<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 143/162.				
ACCESSION	AL021841 AL123456				
VERSION	AL021841.1 GI:3261517				
KEYWORDS					
SOURCE	Mycobacterium tuberculosis.				
ORGANISM	Mycobacterium tuberculosis Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Acidimicrobiales; Corynebacteriaceae; Mycobacteriaceae;				
REFERENCE	Cole,S.T., Brosch,R., Parkhill,J., Garber,T., Churcher,C., Harris,D., Gordon,S.V., Eigmeier,K., Gas,S., Barry III,C.E., Tekaia,F.F., Badcock,K., Basham,D., Brown,D., Chillingworth,I., Connor,R., Davies,R., Devlin,K., Felwell,T., Gentles,S., Hamlin,N., Holroyd,S., Hornsbey,T., Jagels,K., Krogh,A., McLean,J., Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A., Rajandream,M.A., Rogers,J., Rutter,S., Seeger,K., Skelton,S., Squares,S., Sqaes,R., Sulston,J.E., Taylor,K., Whitehead,S. and Barrett,B.G. Deciphering the biology of <i>Mycobacterium tuberculosis</i> from the complete genome sequence Nature 393 (6685) , 537-544 (1998)				
AUTHORS	Erratum:[published erratum appears in Nature 1998 Nov 12;396(6707):190]] 2 (bases 1 to 53662) <i>Parkhill,J.</i> Direct Submission Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk On Jun 27, 1998 this sequence version replaced gi:2894210.				
TITLE	Notes: Details of <i>M. tuberculosis</i> sequencing at the Sanger Centre are available on the World Wide Web. (URL, http://www.sanger.ac.uk/projects/M_tuberculosis/) CDS have been renumbered from the original cosmid submissions but the old gene designations are in brackets after the new gene numbers. Gene prediction was based on a Hidden Markov Model of TB genes implemented in TParse (Krogh) supplemented with visual inspection of positional base preference in codons, especially where there is an increase in the observed/expected third position G + C. CAUTION: In some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.				
JOURNAL					
COMMENT					

FEATURES

Source	Location/Qualifiers
	1..53662
	/organism="Mycobacterium tuberculosis"
	/strain="H37Rv"
	/db_xref="taxon:1773"
	<1..50056
misc_feature	/note="fragment designated v016. Does not represent a physical clone" complement(9..13) /note="possible RBS, aggag, for RV2249c" complement(15..1496) /gene="lpda" complement(15..1496) /gene="Ipda"
RBS	
gene	
CDS	

gene
CDS
/gene="RV3303c", (MTV016_02c), len=493, Probable
dihydrofolate dehydrogenase similar to many e.g.
004829|DLDH_HAIYO (474 aa) from HALOREPRX VOLCANITI. FASTA
scores: sp|004829|DLDH_HAIYO DIHYDROFOLATE
DEHYDROGENASE (E (474 aa) opt:675 z-score: 794.6 E(): 0;
29.3% identity in 471 aa overlap. Tbpase score is 0.883"
/codon_start=1
/transl_table=11
/product="1pda"
/protein_id="CAI17075.1"
/db_xref="GI:2894212"
/db_xref="SPTREMBL:O53355"
/translation="MTWRIVLGGGPGAGYEALVAATSTSHETQVTVIDCGIGAAV
LDDCVSTFTASTGLTFLTFLRRAPHLGHIDFPDAKISLPQIHNRVTTAAAGSADIT
AOLLSGEOVIAVAREGLIDSTPGLAHRHKTADSGSEHADVVAATGASPRILP
SNQPGEMITLMROLYXDLDLPHLYVSGSCVGAPEVDVTELVGVATVVAQDHL
PREDADALVLEESFAERGVLYFKNRAASTTRIGAGVLYTMIDGRVESHMLMTIG
SVYPTSGIGLEKVGIQIGRNVLTVDVRSRLTAGVTAADGCTGLPLASVAAAGGRI
AMAYALGASVLEPILRLRTVAATVFTPRPEIAAGVQSVIDAGSVARTIMPLPTNA
KXSEKRHEFVKIKFCRRSTGVIGGVAAPIASELILPIAAVONRITVAGTOLAVY
PILSGSITEAREALMAHDDLDTCTAAPDAQDALVLPNHLPTSN"
1699.. 2178
/gene="RV3304"
1699.. 2178
/gene="RV3304"
/note="RV3304", (MTV016_03), len= 159, unknown but very
similar to hypothetical protein L308_F3_97 (174 aa) from
M. lepre and some similarity to Q44546 (183 aa) from
Azotobacter vinelandii. ORF continues upstream with
possible start at2198. FASTA scores: sp|tr|Q49927|Q49927
L308_F3_97 (174 aa)opt: 974 z-score: 1017.7 E(): 0; 91.3%
identity in 149 aa overlap and sp|tr|Q44546|Q44546
HYPOTHETICAL 20.3 kD PROTEIN. (183 aa) opt: 206 z-score:
227.5 E(): 3.5e-05; 29.0% identity in 145 aa overlap.
Tbpase score is 0.922"
/codon_start=1
/transl_table=11
/product="hypothetical protein RV3304"
/protein_id="CAI17076.1"
/db_xref="GI:2894213"
/db_xref="SPTREMBL:O53356"
/translation="MPLVAAAGSNMHPOMLERAPSHSPMAGTGLPMRLTFPGEDIG
WEGALIVAEVDEDSKLVFVULDTMPADKKNIIDREGSEFGIHKIRCRVRISSDPTT
DEVLAIVLVAAMEGSLPSARYLGVMAADAIAPSPDYVNDLFTFRANIGCGTIA"
complement(2197.. 3366)
/gene="amaA"
complement(2197.. 3366)
/gene="amaA"
/note="RV3305c", (MTV016_04c), len: 389, amaA, Probable
N-acyl-L-amino acid amidohydrolase or peptidase similar to
many e.g. Bacillus subtilis BSUB0015_194 (416 aa) and
P54964|I1100.SYNY3 (393 aa). Contains P500633 Eukaryotic
L101 (cysteine) proteases histidine active site. FASTA
scores: Z99118|BSUB0015_194 (416 aa) opt: 805 z-score:
953.3 E(): 0; 37.1% identity in 364 aa overlap and
sp|P54964|I1100.SYNY3 HYPOTHETICAL 42.7 kD PROTEIN SL10
(393 aa) opt: 802 z-score:950.2 E(): 0; 38.0% identity in
382 aa overlap. Tbpase score is 0.905"
/codon_start=1
/transl_table=11
/product="amaA"
/protein_id="CAI17077.1"
/db_xref="GI:2894214"
/db_xref="SPTREMBL:O53357"
/translation="MSLDAAESMLAHNDLVGRNRHRIHYPELGRQEVATYQFAE
RLADGILPKVPLVGGTGLTCDFGPDHORIALRLMDMLLPAAETGAPVASTMPNVAN
ACGDHAHRAILGLGALASVPELADVGRLLTFQAAEELMPGADAIADALAGVSTI
FALHGDPLEVKGAVROGPITSADSEILTVPSGCHTSPHLTADLVGLCTLVYTG
LGVSLRLIDRPNSTVLVWAGVNMAGMANAIPQTVGSISRTQVTDWDEELVQ
ALISALLRLAIEHTLYORGVGVAVNEISIRILAAHIAIGIPCVLADTQSGGGEF
SHYLEEVGAAGARUGWSGSDGLQDLNDPPTDIDERLALIGLRVAVIIQAAAH"
complement(2407.. 2439)
/gene="amaA"

CDS join(16052..16064,16146..16319,16355..16935)
 /note="hypothetical protein"
 /codon_start=1
 /protein_id="BA90367.1"
 /db_xref="gi:6815081"
 /translation="MGPTLGAAVEGERGDRMRSGDGGDRGREVAGGAGARVARKGR
 RANGELGOLAAAGEETVGRGERRGACVAAHGAAMPKSGVAPTSLSFEETLNLINLAGI
 ALLAGLIALLAGLPTISFPTSPRASSPKRRSTLEPLCRSYVAATADINDCLARTKR
 EETVYTGRTWRSSPEAAAPITDLRSTPLPRCAVNCFSPLPRVAGVARKGLPASP
 SKSPKLPTRROPPLPSLAHRLHRIATPPAALSH"
 18583..20061
 /note="ESTs C71763(E0309),A0082059(E0309) correspond to a
 region of the predicted gene.
 Similar to Arabidopsis thaliana chromosome II BAC T20B5
 genomic sequence; hypothetical protein. (AC002409)"
 /codon_start=1
 /protein_id="BA90368.1"
 /db_xref="gi:6815082"
 /translation="MAIVAGHCRELSPFGGCVGRHCLRYTOLDCEYFDHDLDD
 RERNHGAGLSPRSLSSPCESTVVPAPADRAEPARDGEMKMPVEVETVAGHAGVH
 ELIAAAGAIATGRLVPAOAGIGGALLLEGRSDVHAATIKPLDDASSSHGGGIG
 YASKVLVEVAATGLDHDFARVAPETALIKTSRPMPTTASIOFPAHECDAGLGP
 SFEVSASVIRGSLDRLPLINDRHAGNLIVKSPSECOASGGSTPLPLDVLIDHGLC
 LPEOLDPVEVLEHWPOSSLSPGSAELIVYASLDSPFRAAMLRALPELTAARILTL
 LCTITLQRAAAGCLLADIDGMTRFEFSAEMEGLSALSCLKNAYDSSTTTAGSPRK
 HHNSDDSDSTQFGMDVPAQLPHLTLGGGGLAKSVSFSAAGGAAGAARAKRM
 SKKALSGDEMAAFLEFQDLPLPALDAKRRAGLKLTRLTGTSF"
 complement(20793..21346)
 /note="3' UTR"
 /join(21752..21851,22552..22676,22886..23101)
 /note="hypothetical protein"
 /codon_start=1
 /protein_id="BA90369.1"
 /db_xref="gi:6815083"
 /translation="MDLKSTSTSVLEKLOEGALPGRTMEREPGGLANSETGRI
 GQMTISGPTTASNVNPVPLCEMNAAREALINPNSGTPARAAHATPHNTAARTAAAP
 AASRSRSMASAPPCGQPPRPMGQSTSTSVAMGCVQCARDAAS"
 complement(join(23394..23529,23538..24018))
 /note="hypothetical protein"
 /codon_start=1
 /protein_id="BA90370.1"
 /db_xref="gi:6815084"
 /translation="MGPRARRKTRRRRRDLAQLRATLSSGRRSGGGGKRGDGRAY
 ECTGAGDCTTKREECTGLYIGSRGLVAGVAATSPAEWRVGRERRDSKIESRPV
 RARTAAEEMARGGRAQAATWATWAPERRGSGRCQMRRLRELVGDPREDEGGR
 LGMVAAFGWPSSAEGGRKGRPRIRKKRREKKIRRK"
 complement(join(24161..24471,24526..24637))
 /note="hypothetical protein"
 /codon_start=1
 /protein_id="BA90371.1"
 /db_xref="gi:6815085"
 /translation="MAVTRVAVHDEGPRTGEGAAVNHVALTWRDVATWQHRTAAGS
 ALRATGSSSRGNSNGVRRVRYGAAGAAAGGERRRDRRATRRRRYRRKREGEE
 GGGVLEFVRPMRETYNGCETROREVDKRTSGTT"
 join(25452..25542,26685..26854,26925..26983,27057..27489
 27514..27825)
 /note="EST AU082058(c12976) corresponds to a region of the
 predicted gene.
 hypothetical protein"
 /codon_start=1
 /protein_id="BA90372.1"
 /db_xref="gi:6815086"
 /translation="MERGGPYDGLCRFRIHLGTGRDCPLPAGDGGPLVDQHAASLK
 EYVAEASPAAAAVDTSGVNPYKRGKRESSVILGRRKTRPSASDASPPRRORLV
 TLGEAAKSAADNRSGNSASPYVAATDVYVNRSEADPNCVGVGLVPGSPAD
 ILTTEELQVEMGRITLQAGRGIRHETIAERAAASASNERADKLADLAEVREDLOKM
 RELVAGNEORGLKLRMSSELENNISLTSRLARECGVTTTIPANDEFLTTTSLA
 ELATAMEVYIPSKHARIGEESSGIYTGTCNVLACVKLHAHELIDLOKILYQGAASNA
 RKMVEIGMGESILPFEF"
 complement(join(29314..30354,33082..35175))
 /note="similar to GAG-POL precursor (AB030283)
 internal stop
 codon; complement(29869..29871,30130..30132,30295..30297,

[illegible]

DEFINITION Equine herpesvirus 4 (EHV4) genome, thymidine kinase (TK) and glycoprotein H (gH) genes.
ACCESSION D14486 D00683 D00684
VERSION D14486.1 GI:221820
KEYWORDS UL21 gene; UL24 gene; gH gene; glycoprotein; thymidine kinase.
SOURCE Equine herpesvirus 4 (strain 1942).
ORGANISM Equine herpesvirus 4
Viruses; dsDNA viruses, no RNA stage; Herpesviridae; Alphaherpesvirinae; Varicelloviruses.
REFERENCE
1 (sites)
Nicolson,L., Cullinane,A.A. and Onions,D.E.
The nucleotide sequence of an equine herpesvirus 4 gene homologue of the herpes simplex virus 1 glycoprotein H gene
J. Gen. Virol. 71 (Pt 8), 1793-1800 (1990)
JOURNAL MEDLINE 90362066
REFERENCE
2 (sites)
Nicolson,L., Cullinane,A.A. and Onions,D.E.
The nucleotide sequence of the equine herpesvirus 4 thymidine kinase gene
J. Gen. Virol. 71 (Pt 8), 1801-1805 (1990)
JOURNAL MEDLINE 90362067
REFERENCE
3 (bases 1 to 5880)
Nicolson,L.
Unpublished (1993)
JOURNAL COMMENT Submitted (10-SEP-1990) to DDBJ by:
Lesley Nicolson
Dept. Veterinary Pathology, University of Glasgow
Vet School
Beardsden Roda, Glasgow G61 1QH, Scotland
UK.
Phone: 041-339-8855
Fax: 041-330-5733.
FEATURES
source
1..5880
/organism="Equine herpesvirus 4"
/db_xref="taxon:10331"
complement(99..104)
/note="put. polyadenylation signal (UL24); putative"
complement(115..120)
/note="put. polyadenylation signal (UL24); putative"
complement(116..934)
/note="putative; ORF1 (UL24 homologue)"
/codon_start=1
/protein_id="BA03377.1"
/db_xref="GI:221821"
/translation="MKRKQRLTARSRLRAGIRCHNRRYNNAMVODLASAKKNVYGARL
APLSELVPAETLKTAMGVSLAEVNLGQRRPDCVTVGFHSGSDAGVCILIELKTC
RFSKNMNTASKNLQKRGMRQLHDSCHRLAKTLPPGSGEILLAPLVFAVQRMRYLR
VTRSLPQVYVSNAAVLSCITSLAEYSPITSESTRRRCVTRRTNSKAFRAKTTGSI
OPTQAKPAATAVAASLFSATQAQNTNAVGYQPATISLANPLAWVASLFAPK"
861..865
/note="put. TATA box (TK); putative"
887
/note="put. TATA box (TK); putative"
misc.feature
/note="pot. transcription initiation site; putative"
949..2007
/partial
/note="ORF2; (TK)"
/codon_start=1
/protein_id="BA03378.1"
/db_xref="GI:221822"
/translation="MAACVPPGEAPRSASGPTTRROYTVIRIYLDGVYGIKSTTGRV
MASAASGSPDLYCEPEPMAYMRTLFETDVISGIVDQNRKQGNLAADADALITAHYO
SRPTPYLLIHDHCTLPFGNSLQORGPOTDLPFDHNPVASTYCFPAARILLDSM
CALMAYATLREPQGNIVYTLINVEHILRLTRARIGQIDITLILATIRNYFMIL
VNTCHFLRSRGVWMDGELPTSCGAYKHRATQDAQERSPGLGTLALFTQEL
LDRCGVILEVHAWMLDALMLKRLNLNVFSADLSGTPQCAAVESLLPLMSLSDSD
SASALERAARFNAEMGV"
2049..2054
/note="put. polyadenylation signal (TK); putative"
2127..2133
/note="putative"
2197..2202
/note="putative"

CDS
2225..4792
/partial
/note="ORF3; (gH)"
/codon_start=1
/protein_id="BA03379.1"
/db_xref="GI:221823"
/translation="MSQPYLKLIALIVATIVSAIPWTTPTVSTSPPOOTKLHYNGCT
WVHNFTNRYDRTIMEPYNNLSITSTFEVIAISENFRVNPVPGASVPLIKSL
NPKRHQPCIANVPERGDPGRGCNASTYSLTFNDNLEFLTKNLEPEVLPDNYTGH
TFERSKTVAIRKNGVGLSPRTSPVNTQTATAYLNGFMGIQISSVNNPLEMLV
TKTLTYISPPAAWPIAFHAGITVMCGDTQATAYLNGFMGIQISSVNNPLEMLV
APNDVRAIRVRLPPRRRLPEPGVAGPIYKVVYLSGNGFYLGHGSKISREVAAYPE
ESLDYRHLISLANDITLAMELSSGSKVSYLYEILARLAVTSLAEVRLSY
MLQEAIDVDINLRIYPLVMKYAAGTADSSTSSVDNADPEVQAQIEKIVADIN
IENLRKPYMEHRSLSYVAYSKRPIPNVSPNRNRTITMAYKEAIDRTWNSTARE
VLPFVGAAGSHVLTIDGPDGLHAKDSMFLSLRNILLCTACTASHAVSAY
KLEEVMAGLIAGVQFSLLEVFSPCMASAFEDLAEHSHVLDLSVIPRLYTDLNTGL
EDDGTTHSGRSANGILNSRIAYNPAVVFPELASCSTKLPRKVLVPLASNRSY
VITRTAPNIGLTYSLDGVNIAKPIVISTYITGVNENQVSRATIRSYLDHPHTOSCYC
GSVEMRYMASGAIYMDLIYIDDKDVELQVAGENSTIPAFNPKTYTPSMNLLMPPNGT
VTLMASAFASAFRIPSTYILMASIGGLIALILYIVKMLCGVINNDYSLINSE"
4777..4782
/note="put. polyadenylation signal (gH); putative"
4893..5315
/note="direct repeat region"
5395..5403
/note="put. replication origin; putative"
5440..5448
/note="put. replication origin; putative"
complement(5755..5880)
/note="putative; ORF4 (UL21 homologue)"
/codon_start=1
/protein_id="BA03380.1"
/db_xref="GI:221824"
/translation="NVSLLTAFDMLDSRGAYLYLYLSERLKGVPVHVNRA"
BASE COUNT 1459 a 1377 c 1499 g 1545 t
ORIGIN
Query Match 4.3%; Score 41.6; DB 59; Length 5880;
Best Local Similarity 44.1%; Pred. No. 25;
Matches 173; Conservative 0; Mismatches 219; Indels 0; Gaps 0;
OY 145 gcgagcggaacacggggtgtgtgtccctggcgacatgacggcgagagagatcatacc 204
DB 4893 GCGGCGGCGAGGGCTGTCGCGCGCGCGAGGGCTGTCGCGCGCGAGGGCTGCGGCGG 4952
OY 205 aatacgaattcgcgcgtgcatccctgcatacccccctgtttcagcgccagccaag 264
DB 4953 GCGAGGCGTGGCGCGCGCGCGCGAGGGCTGCGCGCGCGCGCGAGGGCTGCGGCGG 5012
OY 265 gcaagcgacctcttcgc 324
DB 5013 GCGGCGAGGGCTGCGCGCGCGCGCGAGGGCTGTCGCGCGCGCGCGAGGGCTGCTCG 5072
OY 325 gctaatcagtgatcgaatcgtgggataatgtcaccacaagagagtgctgaagcctgga 384
DB 5073 GCGGCGGCGAGGGCTGTCGCGCGCGCGCGAGGGCTGTCGCGCGCGCGCGAGGGCTGCT 5132
OY 385 ggcgcgcgcacagcagaggtctcaagcggtgacgtgcaacagcaccgcgcgcgcgcgcgc 444
DB 5133 GCGGCGGCGGCGAGGGCTGTCGCGCGCGCGCGAGGGCTGCTGCGCGCGCGCGAGGGCT 5192
OY 445 agtcatacgcgcacaaatgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 504
DB 5193 GCTGCGCGCGCGCGAGGGCTGCTGCTGCGCGCGCGCGAGGGCTGCTGCGCGCGCGG 5252
OY 505 gccagcgcgccctgctccccaagccgcgcgc 536
DB 5253 GCGAGGCTGCTGCGCGCGCGCGAGGGCTGCG 5284

Search completed: September 21, 2001, 00:34:56
Job time: 29551 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 21, 2001, 02:41:41 ; Search time 330.77 Seconds
(without alignments)
1852.744 Million cell updates/sec

Title: US-09-138-735-1_COPY_1232_2207

Perfect score: 976
Sequence: 1 cagtgacagcgtaacgctt.....tgtagacgagcattatgt 976

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_0601:*

1: /cgnl_9/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /cgnl_9/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /cgnl_9/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /cgnl_9/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /cgnl_9/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /cgnl_9/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /cgnl_9/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /cgnl_9/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /cgnl_9/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /cgnl_9/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /cgnl_9/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /cgnl_9/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /cgnl_9/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /cgnl_9/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /cgnl_9/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /cgnl_9/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /cgnl_9/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /cgnl_9/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /cgnl_9/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /cgnl_9/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /cgnl_9/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /cgnl_9/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	976	100.0	3402	20	AAx84092
2	972.8	99.7	3402	17	AAx727310
3	40.4	4.1	477	21	AAx24307
4	40.4	4.1	567	21	AAx29550
5	40	4.1	10732	21	AAx10594
6	39.2	4.0	2790	19	AAx18480
7	38.2	3.9	2943	17	AAx16480
8	38	3.9	1037	21	AAx59242
9	38	3.9	1159	21	AAx59240
10	38	3.9	1472	21	AAx59241
11	37.8	3.9	371	20	AAx19383

12	37.8	3.9	371	20	AAx19171	M. tuberculosis re
13	37.6	3.9	2115	20	AAx32263	Human duncer-like p
14	37.6	3.9	2115	21	AAx88198	PPD39 human duncer
15	37.6	3.9	72750	21	AAx81468	N. meningitidis pa
16	37.6	3.9	349980	21	AAx21544	Neisseria meningit
17	37.2	3.8	921	20	AAx33558	Rice beta-glucanase
18	37.2	3.8	996	20	AAx33558	Rice GUS CDS. Or
19	37.2	3.8	2169	20	AAx33534	Rice beta-glucanase
20	36.8	3.8	3737	21	AAx64656	Partial sequence c
21	36.8	3.8	3776	21	AAx42903	TRP-1 protein codi
22	36.6	3.8	203	19	AAx30271	Glutamine rich reg
23	36.6	3.8	203	19	AAx17226	SCA2 gene CAG repe
24	36.6	3.8	913	17	AAx30253	Cotton fibre cell-
25	36.6	3.8	913	17	AAx13034	Cotton fibre-speci
26	36.6	3.8	913	17	AAx07199	Cotton fibre-speci
27	36.6	3.8	913	18	AAx70041	Cotton fibre-speci
28	36.6	3.8	913	18	AAx62610	Cotton fibre-speci
29	36.6	3.8	913	21	AAx23545	Cotton fibre-speci
30	36.6	3.8	1966	20	AAx61220	CDNA sequence a co
31	36.6	3.8	1984	17	AAx13030	Mouse DNA demethyl
32	36.6	3.8	1985	17	AAx30250	Cotton fibre-speci
33	36.6	3.8	1985	18	AAx70036	Cotton fibre clone
34	36.6	3.8	3324	20	AAx60262	Cotton H6 gene and
35	36.4	3.7	2415	18	AAx6757	Nucleic acid sequ
36	36.4	3.7	2453	18	AAx6758	CDNA of the M3/6 g
37	36.2	3.7	336	20	AAx88093	Genomic antifer
38	36.2	3.7	633	11	AAx03322	Synthetic Elmeria
39	36.2	3.7	862	21	AAx09682	Fusarium venenatu
40	36.2	3.7	3201	21	AAx14955	DNA encoding a PDE
41	36	3.7	795	19	AAx55830	FLGA insert stabl
42	36	3.7	1448	21	AAx55831	Nucleotide sequen
43	36	3.7	1448	21	AAx21931	Human breast and o
44	36	3.7	1925	20	AAx0924	Epstein Barr Virus
45	36	3.7	1926	21	AAx50254	Epstein Barr Virus

ALIGNMENTS

RESULT 1

ID AAx84092 standard; cDNA: 3402 BP.

AAx84092:

27-AUG-1999 (first entry)

T. cruzi PTC40 coding sequence.

PTC40: Tc40: infection; diagnosis; immune complex; antigenic determinant; therapy; antibody; ds.

Trypanosoma cruzi.

WO9929867-A1.

17-JUN-1999.

10-DEC-1998; 98WO-1B01987.

10-DEC-1997; 97US-0988242.

(INMR) BIO MERIEUX.

Jolivet M, Lesenechal M, Mandrand B, Paranhos-Baccala G;

WPI: 1999-394978/33.

P-PSDB: AAY2124.

New Trypanosoma cruzi antigen

Claim 1; Page 52-56; 65pp; English.

CC This sequence encodes the Trypanosoma cruzi Prc40 protein of the
CC invention, and is designated Prc40. The Prc40 antigenic determinant is
CC useful as a reagent for detection and/or monitoring of Trypanosoma cruzi
CC infection from samples including blood serum or plasma, urine, saliva, or
CC tears, by contacting with the sample and detecting an immune complex.
CC Prc40 antigenic determinant, the vector, expression cassette, cell or
CC antibody are useful for treatment or prevention (vaccine) of a
CC Trypanosoma cruzi infection in a man or animal. Current Trypanosoma cruzi
CC antigens are obtained from protein fractions of the noninfectious stage
CC of the parasite, and these do not allow sufficient production of antigens
CC for use in reliable serological diagnostic tests. The strain to strain
CC polymorphism reduces reliability of the tests.

SQ Sequence 3402 BP; 888 A; 821 C; 956 G; 737 T; 0 other;

Query Match	100.0%	Score 976;	DB 20	Length 3402;
Best Local Similarity	100.0%;	Pred. No. 2.2e-257;		
Matches 976; Conservative	0;	Mismatches	0;	Gaps 0;

[illegible]

Db	2012	caacttctgaa	ttcgggttt	taagagc	acagatg	acagagctg	ccagcag	ccgtcca	aca	2071
Oy	841	ccgcaaacac	agccggaag	acacaaagc	ctccggaat	tcatcgctg	tttgaag	cgtacac		900
Db	2072	ccgcaaacac	agccggaag	acacaaagc	ctccggaat	tcatcgctg	tttgaag	cgtacac		2131
Oy	901	ctgtcttc	attgtcgga	ttccctctc	tcgcaac	atcacgaag	gggtgaag	cgtgtcg		960
Db	2132	ctgtcttc	attgtcgga	ttccctctc	tcgcaac	atcacgaag	gggtgaag	cgtgtcg		2191
Oy	961	aacgagc	catatgt	atgt	976					
Db	2192	aacgagc	catatgt	atgt	2207					

RESULT	2
AAT27310	
ID	AAT27310 standard; cDNA; 3402 BP.

AC	AAT27310;	
XX		
DT	26-NOV-1996	(first entry)
XX		
DE	Trypanosoma cruzi epimastigotic Prc100t antigen gene.	
XX		
KW	Antigen: Trypanosoma cruzi; epimastigote; serum; Chagas disease; probe	
KW	primer; PCR; polymerase chain reaction; amplification; antibody; ds.	
XX		
OS	Trypanosoma cruzi.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	266..3013
FT		/*tag= a
FT		/product= Prc100t epimastigotic antigen
XX		
PN	FR2723589-A1.	
XX		
PD	16-FEB-1996.	
XX		
PF	12-AUG-1994;	94FR-0010132.
XX		
PR	12-AUG-1994;	94FR-0010132.
XX		
PA	(INMR) BIO MERIEUX.	
PI	Jolivet M., Lesenechal M., Paranhos-Baccala G;	
XX		
DR	WPI: 1996-190287/20.	
XX		
DR	P-PSDB; AAR91615.	
XX		
PT	New nucleic acid encoding Trypanosoma cruzi epimastigotic antigen -	
XX	useful for diagnosis, monitoring and therapy of Chagas disease	
PS	Claim 1; Page 24-26; 55pp; French.	
XX		
CC	This is the nucleotide sequence encoding a novel isolated antigenic	
CC	protein from Trypanosoma cruzi epimastigotes, designated prc100t.	
CC	The clone Tc50 was isolated from a T. cruzi genomic expression library ;	
CC	lambda gt11, using a mixture of sera from patients with Chagas disease	
CC	Clone Tc50 contained an 594 bp insert corresp. to nucleotides 1232-1822	
CC	of this sequence. The Tc50 sequence was subsequently used to probe a	
CC	Southern blot of restriction enzyme digested T. cruzi DNA and also screened	
CC	a lambda gt10 library to isolate a 1041 bp EcoRI fragment corresp. to	
CC	nucleotides 1403-2443 of Prc100t. Primers (AAT27311-5) were synthesized	
CC	based on the sequences of the 594 and 1041 bp fragments and used to	
CC	amplify the Prc100t clone as 3 fragments from cDNA derived from mRNA	
CC	purified from T. cruzi epimastigotes. The protein or antibodies raised	
CC	against it can be used in the detection and monitoring of T. cruzi	
CC	infection i.e. Chagas disease.	
XX		
SO	Sequence 3402 BP; 889 A; 818 C; 958 G; 737 T; 0 other;	

Query Match 99.7%; Score 972.8; DB 17; Length 3402;
 Best Local Similarity 99.8%; Pred. No. 1.7e-256;
 Matches 974; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 caggtacagcgttaacgagcttctgcttcaatcgttacagcgcagcgtagctgcttgcgt 60
    |||
DB 1232 caggtacagcgttaacgagcttctgcttcaatcgttacagcgcagcgtagctgcttgcgt 1291
QY 61 gacatgtcgtatcgtatgaacatcttccatctcgcgtctcccgagggagaagaacagcag 120
    |||
DB 1292 gacatgtcgtatcgtatgaacatcttccatctcgcgtctcccgagggagaagaacagcag 1351
QY 121 ccaagcccaaaaacacgtgtagtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 180
    |||
DB 1352 ccaagcccaaaaacacgtgtagtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1411
QY 181 gacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 240
    |||
DB 1412 gacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1471
QY 241 cccctgttcaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 300
    |||
DB 1472 cccctgttcaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1531
QY 301 ccgcacgtggtgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 360
    |||
DB 1532 ccgcacgtggtgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1591
QY 361 caaaggaagcgtcgtacgacttgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 420
    |||
DB 1592 caaaggaagcgtcgtacgacttgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1651
QY 421 agcgtacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 480
    |||
DB 1652 agcgtacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1711
QY 481 ggtatgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 540
    |||
DB 1712 ggtatgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1771
QY 541 ccaacaggaagaagaagagcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 600
    |||
DB 1772 ccaacaggaagaagaagagcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1831
QY 601 aacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 660
    |||
DB 1832 aacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1891
QY 661 gcaagcagatataccagctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 720
    |||
DB 1892 gcaagcagatataccagctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1951
QY 721 ataaccaacacgcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 780
    |||
DB 1952 ataaccaacacgcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 2011
QY 781 caactctcgtatctggtgttgagagcagatgacagagctcgcgcgcgcgcgcgcgcgcgcgcgc 840
    |||
DB 2012 caactctcgtatctggtgttgagagcagatgacagagctcgcgcgcgcgcgcgcgcgcgcgcgc 2071
QY 841 ccgcaacaacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 900
    |||
DB 2072 ccgcaacaacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 2131
QY 901 ctgtctcatttgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 960
    |||
DB 2132 ctgtctcatttgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 2191
QY 961 aacgagagccattatgt 976
    |||
DB 2192 aacgagagccattatgt 2207
  
```

RESULT 3
 AA244307/c
 ID AA244307 standard; DNA; 477 BP.

AA244307;
 04-APR-2000 (first entry)

Human SCA7 genomic DNA fragment 1.

SCA7; human; spinocerebellar ataxia type 7; SCA1; SCA2; SCA3; SCA6;
 repeat expansion detection; RED analysis; detection; ds.

Homo sapiens.

CA2245310-A.

19-FEB-1999.

19-AUG-1998; 98CA-2245310.

19-AUG-1997; 97US-0056170.

(MINU) UNIV MINNESOTA.

Koob MD, Rannum LP;

WPI: 2000-098181/09.

Identifying individuals at risk of developing spinocerebellar ataxia
 type 7 by analyzing trinucleotide repeat regions of spinocerebellar
 ataxia type 7 gene -

Claim 36; Page 41; 66pp; English.

This invention describes a novel method for identifying individuals at
 risk for developing spinocerebellar ataxia type 7 (SCA7). The method
 comprises analyzing the CAG repeat region of a SCA7 gene to detect CAG
 repeats, where individuals at risk have at least 30 CAG repeats and
 those not at risk have less than 19 CAG repeats. The method is useful
 for identifying individuals at risk of developing SCA7 and also those
 at risk of developing SCA1, 2, 3 or 6. The use of genomic DNA in the
 CC repeat expansion detection (RED) analysis allows isolation of any
 potential trinucleotide repeat expansion regardless of the expression
 pattern. Utilization of different oligonucleotides in the RED assay
 allows any of the possible trinucleotide repeats to be detected, and
 the cycled nature of the reaction makes it extremely sensitive. This
 CC sequence represents a genomic DNA fragment of the human SCA7 gene
 which is described in the method of the invention.

Sequence 477 BP; 99 A; 164 C; 172 G; 42 T; 0 other;

Query Match 4.1%; Score 40.4; DB 21; Length 477;
 Best Local Similarity 53.9%; Pred. No. 0.11;
 Matches 83; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

```

QY 145 ggcagcgcgaacccggggtgtgtctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 204
    |||
DB 334 ggcagcgcgcgaacccggggtgtgtctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 275
QY 205 aataagacttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 264
    |||
DB 274 gctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 215
QY 265 ggcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 298
    |||
DB 214 gctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 181
  
```

RESULT 4
 AAA29550/c

ID	AAA29550	standard; DNA; 567 BP.
XX		
AC	AAA29550;	
XX		
DT	09-AUG-2000	(first entry)
XX		
DE	HIV codon altered env nucleotide sequence #2.	
XX		
KW	Erythropoietin; EPO; G-CSF; granulocyte colony stimulating factor;	
KM	wobble; codon altered gene; shuffling; modification; vaccine;	
KW	insulin; peptide hormone; growth factor; cytokine; interferon;	
KM	leukotkin; leukaemia inhibitory factor; oncostatin M;	
XX	transcription activator; expression activator; infectious organism; ds.	
OS	Human immunodeficiency virus type 1.	
SS	Synthetic.	
PN	WO200018906-A2.	
PD		
XX	06-APR-2000.	
PE		
XX	28-SEP-1999; 99WO-US22588.	
PR	29-SEP-1998; 98US-0102362.	
XX	29-JAN-1999; 99US-0117729.	
PR	05-FEB-1999; 99US-0118813.	
XX	24-JUN-1999; 99US-0141049.	
PA	(MAXY-) MAXGEN INC.	
PI	Patten PA, Liu L, Stemmer WPC;	
DR	WPI; 2000-303449/26.	
PT	Novel methods for recombining codon-altered libraries of nucleic acids	
XX	used to produce new proteins and new vectors with reduced rates of	
PT	reversion to wild type -	
XX		
PS	Example; Fig 18A; 92pp; English.	
CC	A method has been developed of making codon altered nucleic acids (NAs),	
CC	comprising providing a NA sequence (NA1) which encodes a polypeptide	
CC	(P1), providing codon altered NA sequences, each encoding P1 or a	
CC	modified form of it, and recombining the codon altered NA sequences to	
CC	produce a target codon altered NA which encodes a second protein.	
CC	The method of the invention can be used for recombining codon-altered	
CC	libraries of nucleic acids to produce new proteins, which have	
CC	improvements in a desirable characteristic. Target nucleic acids	
CC	include those coding for therapeutic proteins such as erythropoietin	
CC	(EPO), insulin, peptide hormones, growth factors, cytokines, interferons,	
CC	interleukins, leukemia inhibitory factor, and oncostatin M, as well as	
CC	expression activators and proteins from infectious	
CC	organisms for use as vaccines. The method can also be used to produce	
CC	attenuated viruses which have reduced rates of reversion to wild type.	
CC	The present sequence represents an HIV codon altered env nucleotide	
CC	sequence, which is used in an example from the present invention.	
XX		
SO	Sequence 567 BP; 32 A; 129 C; 222 G; 184 T; 0 other:	
	Query Match	4.1%; Score 40.4; DB 21; Length 567;
	Best Local Similarity	49.1%; Pred. No. 0.12;
	Matches 107; Conservative 0; Mismatches 111; Indels 0; Gaps 0;	
Oy	386 cgcgcgcaagcgaagtctacgcggtgaagtcacgaactacgccccgcgaagaaca	445
Dd	477 CGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA	418
Oy	446 gtccataaggccaatbtggcgaacctgtgaagcgttgatgttgcgaactaatagtg	505
Dd	417 CGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA	358
Oy	506 ccagcgcgacctcgctccccaacagccgcgcgcgaacaaacagagagaagaaaagctccg	565

[illegible]

Db 9663 rysaaavathrgaansrstrghrlythrgsgrsthrhggyysgivarashsrngnth 97222

QY 392 ccacgacgaggtctacgcyggtacgtccacgactaccccgacgcgaacaagtcacat 451

Db 9723 rglthrtrgvaaasamtgmtrtaashsrassysrnyaanthrcysrashhsygt 97822

QY 452 accggcccaatgccgaccctgtacgcgcgttggtgacgtatattggtgccacg 511

Db 9783 rtaargytrgnasvavaystlrgnmasnshvatlrhssysyatasashssyvaaaarhs 98422

QY 512 cggcgctcgtctccacagccgcgcgcgaaccca-acagaggaagaaagcctccgcgcga 570

Db 9843 vaasnaashashassrstrargystrvashastlrgnasntrngcyslscysrcysr 99022

QY 571 tbtgaaacgagctccgtgcgcgataaatcgacacgcgcgcgcg 612

Db 9903 asstrsvavaasnrstraaygaanaaasrlyatargtrgng 9944

RESULT	6
AAV18480	
ID	AAV18480 standard; cDNA to mRNA; 2790 BP.

AC	AAV18480;
XX	
DT	18-AUG-1998 (first entry)

KW Corticotroph; pituitary; BOP1; tumour suppressor; constitutive;
 KW inducible; alzheimer's disease; nuclear transcription factor; apoptosis;
 KW cell cycle; neuronal disorder; ss.

Mus sp.

	key	Location/Qualifiers
PH		
FT	CDS	542..2545
FT		/*tag= a
FT		/product= "BOP1 protein"

PN W09813489-A1.

PD 02-APR-1998.

PF 22-SEP-1997; 97WO-EP05198.
XX

PR 23-SEP-1996; 96US-0718661.

PA (CNRS) CENT NAT RECH SCI.
PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN

PI Journot L, Spengler D;
XY

DR WPI: 1998-230701/20.

DK P-PSDB; AAW48760.
XY

PT New isolated tumour suppressor gene - useful for developing products
PT for use in diagnosis and treatment of tumour(s) or neuronal
PT disorder(s)
VV

PS Claim 1; Pages 72-76; 118pp; English.

CC The present sequence represents the Bop1 cDNA isolated from the
CC mouse corticotroph pituitary tumour cell line AtT-20 cDNA library.
CC The protein encoded by the Bop1 cDNA displays a tumour suppressing
CC activity when it was constitutively and inducibly expressed in
CC tumour cells. The Bop1 cDNA and the protein it encodes are claimed
CC to be useful in the preparation of therapeutic compositions, useful
CC for treating, preventing or delaying the recurrence of a tumour or
CC neuronal disorders, e.g. genetic diseases or acquired degenerative
CC encephalopathies such as Alzheimer's disease. The Bop1 protein is
CC also claimed to be able to induce apoptosis resulting in inhibition
CC of tumour cell growth, to suppress tumour formation, to induce G1

CC	arrest of the cell cycle and to act as nuclear transcription factor.
XX	
sq	Sequence 2790 BP; 667 A; 783 C; 714 G; 626 T; 0 other;

Query Match	4.08;	Score 39.2;	DB 19;	Length 2790;
Best Local Similarity	54.18;	Pred. No. 0.45;		
Matches 80; Conservative	0;	Mismatches 68;	Indels 0;	Gaps 0;

QY 137 cggtagtgcgacgcgaaacgggtgtgttcctcgcgcactgcgcgcagtagca 196
| | | | | | | | | | | | | | | | | |
Db 2114 cagcagcagccaccgcgcactcgcgcgcgcaccaccgcgcactgcacatgccacagcca 2173

QY 197 gtactaccaatagactcttgcgcgctgtcatccctgcataccccctgtttcaggcg 256
|| || | ||| | ||| | | ||| | | | ||| |
Db 2174 gtgccacagccaactaccacagcgccacagatgcagccacagtttcagtgcagatccagccc 2233

Qy	257	cagccaagcgcgacgcgcctctctgcgc	284
Db	2234	cagatgcagctaccacagctgtgcgc	2261

RESULT	7
AAT16480/c	
ID	AAT16480 standard; DNA; 2943 BP.

DT	11-MAY-1996	(first entry)
XX		
AC	AA16480;	
.....		

KW Herpes simian monkey SA8 virus gB glycoprotein; immunoassay;
KW diagnosis; herpes B virus; ss.

OS Herpes simian monkey SA8 virus.

Key	Location/Qualifiers
FH	87..2744
ET	/*tag= a
FT	/product= gB_glycoprotein
FT	

PN US5487969-A.

PD 30-JAN-1996.

PF 01-APR-1993; 93US-0042747

PR 01-APR-1993; 93US-0042747

PA (SWBI -) SOUTHWEST FOUND BIOMEDICAL RES.

PI Black D, Eberle R, Hilliard J, Scnicariello F
XX

DR WPI; 1996-105220/11.
DR P-PSDB; AAR92747.

Detection of harr

PT Detection of herpes B virus by PCR amplification of sample DNA - to
PT detect a specific herpes simian monkey B virus DNA segment.

PS Disclosure; Column 23-30; 22pp; English

CC This sequence encodes the herpes simian monkey SA8 virus gB
CC glycoprotein. Such sequences have been used to develop synthetic
CC DNA primers which have homologous sequences of conserved regions
CC which flank a divergent region of the gB glycoprotein gene. In
CC addition, proteins such as the gB glycoprotein have potential use in
CC the development of serological immunoassays. One approach is to
CC synthesize peptides which, based on the properties of the predicted
CC protein sequence, are likely to be immunologically active. Such
CC peptides can be used as substrate antigens in immunoassays to detect
CC serum antibodies which recognize this specific peptide sequence.
CC Synthetic peptides may also be used to produce antibodies against

b6
c7C

ctgcccccgccgcacccctcgcccagccccaagtgtcaagccgccgccgg 636

XX
Sequence 72750 BP. 17518 A: 19945 C: 18810 G: 16477 T: 0 other.

SQ Sequence 72750 BP; 17518 A; 19945 C; 18810 G; 16477 T; 0 other;

117: qb_est48:*
118: qb_est49:*
119: qb_est50:*
120: qb_est51:*
121: qb_est52:*
122: qb_est53:*
123: qb_est54:*
124: qb_est55:*
125: qb_est56:*
126: qb_est57:*
127: qb_est58:*
128: qb_est59:*
129: qb_est60:*
130: qb_est61:*
131: qb_est62:*
132: qb_est63:*
133: qb_est64:*
134: qb_est65:*
135: qb_est66:*
136: qb_est67:*
137: qb_est68:*
138: qb_est69:*
139: qb_est70:*
140: qb_est71:*
141: qb_est72:*
142: qb_est73:*
143: qb_est74:*
144: qb_est75:*
145: qb_est76:*
146: qb_est77:*
147: qb_est78:*
148: qb_est79:*
149: qb_est80:*
150: qb_est81:*
151: qb_est82:*
152: qb_est83:*
153: qb_est84:*
154: qb_est85:*
155: qb_est86:*
156: qb_est87:*
157: qb_est88:*
158: qb_est89:*
159: qb_est90:*
160: qb_est91:*
161: qb_est92:*
162: qb_est93:*
163: qb_est94:*
164: qb_est95:*
165: qb_est96:*
166: qb_est97:*
167: qb_est98:*
168: qb_est99:*
169: qb_est100:*
170: qb_est101:*
171: qb_est102:*
172: qb_est103:*
173: qb_est104:*
174: qb_est105:*
175: qb_est106:*
176: qb_est107:*
177: qb_est108:*
178: qb_est109:*

190: qb_est110:*
191: qb_est111:*
192: qb_est112:*
193: qb_est113:*
194: qb_est114:*
195: qb_est115:*
196: qb_est116:*
197: qb_est117:*
198: qb_est118:*
199: qb_est119:*
200: qb_est120:*
201: qb_est121:*
202: qb_est122:*
203: qb_est123:*
204: qb_est124:*
205: qb_est125:*
206: qb_est126:*
207: qb_est127:*
208: qb_est128:*
209: qb_est129:*
210: qb_est130:*
211: qb_est131:*
212: qb_est132:*
213: qb_est133:*
214: qb_est134:*
215: qb_est135:*
216: qb_est136:*
217: qb_est137:*
218: qb_est138:*
219: qb_est139:*
220: qb_est140:*
221: qb_est141:*
222: qb_est142:*
223: qb_est143:*
224: qb_est144:*
225: qb_est145:*
226: qb_est146:*
227: qb_est147:*
228: qb_est148:*
229: qb_est149:*
230: qb_est150:*
231: qb_est151:*
232: qb_est152:*
233: qb_est153:*
234: qb_est154:*
235: qb_est155:*
236: qb_est156:*
237: qb_est157:*
238: qb_est158:*
239: qb_est159:*
240: qb_est160:*
241: qb_est161:*
242: qb_est162:*
243: qb_est163:*
244: qb_est164:*
245: qb_est165:*
246: qb_est166:*
247: qb_est167:*
248: qb_est168:*
249: qb_est169:*
250: qb_est170:*
251: qb_est171:*
252: qb_est172:*
253: qb_est173:*
254: qb_est174:*
255: qb_est175:*
256: qb_est176:*
257: qb_est177:*
258: qb_est178:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :

COMMENT	FEATURES	BASE COUNT	ORIGIN
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammosser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RcCl-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm .	source 1. 925 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone_lib="RcCl-98" /clone="BACR19D16" /note="end : TET3"	120 a 61 c 61 g 172 t 511 others	
Query Match Best Local Similarity 14.0%; Pred. No. 0.0018; Matches 55; Conservative 172; Mismatches 166; Indels 0; Gaps 0;		5.4%; Score 52.6; DB 219; Length 925; 14.0%; Pred. No. 0.0018; 172; Mismatches 166; Indels 0; Gaps 0;	
221 ctgcgcgcatccctcgatcaccccccgttgcggcccaagaagcgcgcctctg 280			
925 GSBSCSCSCGSCSBSCSSSMSTSSNSBSCSSBSSSTSSKSSBSSSSSS 866			
281 ccgc 340			
865 SSGTSSACVKCANASSSCCGCGGMAHCMCSSSSSCGASARGVARSGAGKRGGS 806			
341 agctgggagtaattgtaccacaagaagcgttcgtcagcacttgagcgcgcgcga 400			
805 GGASASHSSSSACBSSSSSCASCMWASSSSSSASRSRSGGAGGASASSKSSSS 746			
401 ggtctacgcgcgtgcacgcactacacgcgcgcgcgcgcgcgcgcgcgcgc 460			
745 ASAGSVSSASSSSSSSCSSSVSCSSVASSMCCSBSSSSASASSSSSSASCA 686			
461 atcgccgcaacttgacgcgttgattgtgtgcacgtataatgtgtgtgcacgcgcgc 520			

QY	521	ctccacacacccgcggcgcaaacacacagagagaagaagccctccgcggcatttgaaga	580
	522		581
	523		582
	524		583
	525		584
	526		585
	527		586
	528		587
	529		588
	530		589
	531		590
	532		591
	533		592
	534		593
	535		594
	536		595
	537		596
	538		597
	539		598
	540		599
	541		600
	542		601
	543		602
	544		603
	545		604
	546		605
	547		606
	548		607
	549		608
	550		609
	551		610
	552		611
	553		612
	554		613
	555		614
	556		615
	557		616
	558		617
	559		618
	560		619
	561		620
	562		621
	563		622
	564		623
	565		624
	566		625
	567		626
	568		627
	569		628
	570		629
	571		630
	572		631
	573		632
	574		633
	575		634
	576		635
	577		636
	578		637
	579		638
	580		639
	581		640
	582		641
	583		642
	584		643
	585		644
	586		645
	587		646
	588		647
	589		648
	590		649
	591		650
	592		651
	593		652
	594		653
	595		654
	596		655
	597		656
	598		657
	599		658
	600		659
	601		660
	602		661
	603		662
	604		663
	605		664
	606		665
	607		666
	608		667
	609		668
	610		669
	611		670
	612		671
	613		672
	614		673
	615		674
	616		675
	617		676
	618		677
	619		678
	620		679
	621		680
	622		681
	623		682
	624		683
	625		684
	626		685
	627		

AUTHORS Cordonnier-Pratt, M.-M., Gingle, A., Marsala, C., Sudman, M. and F. H.

- Web : www.genoscope.cns.fr

COMMENT
Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) -


```
/cultivar="Blanco"
/db_xref="taxon:4550"
/clone="WHE0517_A09_A17"
/clone_lib="Secale cereale aluminum-stressed root tip cDNA
library"
/tissue_type="Root tip"
/dev_stage="Seedling"
/lab_host="E. coli DH12S"
/note="Vector: pSPORI1; Site_1: Salt; Site_2: NotI; Seeds
were germinated aseptically on filter paper and
transferred to a hydroponic growth system in a growth
chamber when the primary root was 1 cm in length. After a
2-day establishment period, seedlings were subjected to a
5 ppm aluminum stress prior to tissue harvest. Plants
were grown in an environmental chamber. The tissue, total
RNA, and poly(A) RNA were prepared, and a cDNA library
was made (Butler and Gustafson) at University of Missouri
, Columbia. Plasmid DNA preparations and DNA sequencing
were performed in the OD Anderson lab (all other authors
)."
BASE COUNT      97 a      151 c      205 g      124 t
ORIGIN
```

```
Query Match      4.4%; Score 43; DB 137; Length 577;
Best Local Similarity 48.6%; Pred. No. 0.61;
Matches 118; Conservative 0; Mismatches 125; Indels 0; Gaps 0;
```

```
QY 470 cgtgaacgctgattgtgacagataatggtgcagcgccgtctccacag 529
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 19 CAGCTACGAGGTGTGTGCGGACGCTACGGCGCGCGCGCGGCTACGGCTACGGCGG 78

QY 530 ccgcgaggaacacacagaggaagaagccctccgcgcatgtgaacgagctctg 589
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 79 CGGCGCGGAGCGCTACGAGAGAGAGAGCGCGCGCGCTACGCTACGCTACGCTACG 138

QY 590 cgataatgcagacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 649
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 139 GGAAGGCGCATCTCCAGGAGACTGCACCGAGCGCGCGCGCGCGCTGTGAGAG 198

QY 650 ggcgttgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 709
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 199 CGGCGGTGCGCGCTGCGCGCGCGCGCGCGCGCGCGCTGCTTCGCTACGCGAGTC 258

QY 710 tgg 712
      |||
Db 259 TGG 261
```

```
RESULT 9
CNS0091P      925 bp      DNA      GSS      03-JUN-1999
LOCUS      Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION      BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION      AL053013
VERSION      AL053013.1 GI:4934461
KEYWORDS      GSS.
SOURCE      fruit fly.
ORGANISM      Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachyera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 925)
Genoscope.
REFERENCE      Direct Submission
AUTHORS      Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
TITLE      BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
JOURNAL      - Web : www.genoscope.cns.fr)
COMMENT      Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
```

please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazuhiro Osoegawa and Aaron Mammosier in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; on bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

```
FEATURES
source
1. 925
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR19D16"
/note="end : TET3"
BASE COUNT      120 a      61 c      61 g      172 t      511 others
ORIGIN
```

```
Query Match      4.4%; Score 42.8; DB 219; Length 925;
Best Local Similarity 12.8%; Pred. No. 0.73;
Matches 45; Conservative 154; Mismatches 153; Indels 0; Gaps 0;
```

```
QY 369 cgtgtcagcactgagcgccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 428
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 555 SSGGSGYKGGSGSGGSGSCSSSCSSSCSCBCCCCSSSYCCSSSHSSSKSSTS 614

QY 429 cgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 488
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 615 BSCGCCSSKSVCTSCSSSSSSSSSSSTSSSTSSSTSSSTSSSTSSSTSSSTSS 674

QY 489 ggcgcataatggtgcccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 548
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 675 TSAGSGSWAGGSGSGTGTSSSSSSSSSTSSSSSTSSSSSTSSSGKSTBSGSSSS 734

QY 549 agagaagaagcctccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 608
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 735 STSSBBSCTSTSSSSSSSSSTSSCTCCCTCSYSSSTSSSTSSSTSSSTSSSVGT 794

QY 609 ggcgttcacacgcctctccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 668
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 795 SSSSDSTSTCCSCCTTCCTBTBMCYTSTSCGSGSSSGKGVTKCGCGGSSSTN 854

QY 669 ataccagtcgagcgaggttcacagtcgctgagcgagtcgagtcgctc 720
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 855 GMBGTSSACSSSSSSSVSSSSKSSASSSSVSSSGSSGVSSSSASKS 906
```

```
RESULT 10
CNS010B7      787 bp      DNA      GSS      26-JUL-1999
LOCUS      Drosophila melanogaster genome survey sequence T7 end of BAC
DEFINITION      BACN03M07 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION      AL098749
VERSION      AL098749.1 GI:5610360
KEYWORDS      GSS.
SOURCE      fruit fly.
ORGANISM      Plasmid Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachyera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 787)
Genoscope.
REFERENCE      Direct Submission
AUTHORS      Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
TITLE      BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
JOURNAL      - Web : www.genoscope.cns.fr)
COMMENT      Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 21, 2001, 02:16:14 : Search time 164.23 Seconds
(without alignments)
1125.054 Million cell updates/sec

Title: US-09-138-735-1_COPY_1232_2207

Perfect score: 976

Sequence: 1 cagtgacgcgcgtacgcgtcgttgcgttcacatcgtacgcgcgttagctgctcgtcgt

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA: *
1: /cgnl_7/ptodata/1/ina/5A.COMB.seq: *
2: /cgnl_7/ptodata/1/ina/5B.COMB.seq: *
3: /cgnl_7/ptodata/1/ina/6A.COMB.seq: *
4: /cgnl_7/ptodata/1/ina/6B.COMB.seq: *
5: /cgnl_7/ptodata/1/ina/PCPUS.COMB.seq: *
6: /cgnl_7/ptodata/1/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	976	100.0	3402	1 US-08-480-917-1	Sequence 1, Appl
2	39.2	4.0	2790	2 US-08-718-661-1	Sequence 1, Appl
3	38.2	3.9	2943	1 US-08-042-747A-7	Sequence 7, Appl
4	37.6	3.9	2115	2 US-08-474-379C-60	Sequence 60, Appl
5	37.6	3.9	2115	3 US-09-146-249A-60	Sequence 60, Appl
6	37.6	3.9	2115	3 US-08-206-188B-60	Sequence 60, Appl
7	36.8	3.8	3777	3 US-09-121-321-15	Sequence 15, Appl
8	36.8	3.8	3777	4 US-08-933-803A-15	Sequence 15, Appl
9	36.6	3.8	203	4 US-09-043-303-7	Sequence 7, Appl
10	36.6	3.8	913	1 US-08-217-327-3	Sequence 3, Appl
11	36.6	3.8	913	1 US-07-885-970A-3	Sequence 3, Appl
12	36.6	3.8	913	1 US-08-298-687A-3	Sequence 3, Appl
13	36.6	3.8	913	1 US-08-530-797-2	Sequence 2, Appl
14	36.6	3.8	913	1 US-08-298-829-3	Sequence 3, Appl
15	36.6	3.8	913	2 US-08-787-335-2	Sequence 2, Appl
16	36.6	3.8	1984	1 US-07-885-970A-25	Sequence 25, Appl
17	36.6	3.8	1985	1 US-08-298-687A-25	Sequence 25, Appl
18	36.6	3.8	1985	1 US-08-298-829-25	Sequence 25, Appl
19	36.2	3.7	336	2 US-07-814-220-4	Sequence 4, Appl
20	36.2	3.7	336	2 US-07-814-221-4	Sequence 4, Appl
21	36	3.7	2580	3 US-09-050-863-2	Sequence 2, Appl
22	36	3.7	5452	2 US-09-130-114-1	Sequence 1, Appl
23	36	3.7	9600	4 US-08-910-647-1	Sequence 1, Appl
24	36	3.7	10596	1 US-07-884-811-15	Sequence 15, Appl
25	36	3.7	10596	1 US-07-885-971-15	Sequence 15, Appl
26	36	3.7	10596	1 US-08-087-783A-15	Sequence 15, Appl
27	36	3.7	10596	1 US-08-194-088B-15	Sequence 15, Appl

C 28	36	3.7	10596	2	US-08-194-087-15	Sequence 15, Appl
C 29	36	3.7	10596	5	PCT-US93-04648-15	Sequence 15, Appl
C 30	35.2	3.6	13976	1	US-08-320-559-29	Sequence 29, Appl
C 31	35.2	3.6	3376	3	US-08-345-860D-29	Sequence 29, Appl
C 32	35.2	3.6	3376	3	PCT-US94-04496-29	Sequence 29, Appl
C 33	35	3.6	168	1	US-08-469-802B-4	Sequence 4, Appl
C 34	35	3.6	168	1	US-08-469-802B-5	Sequence 4, Appl
C 35	35	3.6	171	1	US-08-469-802B-3	Sequence 5, Appl
C 36	35	3.6	171	2	US-08-267-803B-5	Sequence 5, Appl
C 37	35	3.6	195	1	US-08-469-802B-2	Sequence 2, Appl
C 38	35	3.6	195	2	US-08-267-803B-2	Sequence 2, Appl
C 39	35	3.6	234	1	US-08-469-802B-4	Sequence 3, Appl
C 40	35	3.6	234	2	US-08-267-803B-3	Sequence 3, Appl
C 41	35	3.6	336	2	US-07-814-220-3	Sequence 3, Appl
C 42	35	3.6	336	2	US-07-812-421-3	Sequence 3, Appl
C 43	34.8	3.6	1299	1	US-07-661-610C-11	Sequence 11, Appl
C 44	34.8	3.6	2001	1	US-08-674-168-24	Sequence 24, Appl
C 45	34.8	3.6	2001	3	US-08-985-908-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1
US-08-480-917-1
Sequence 1, Application US/08480917
Patent No. 5820864
GENERAL INFORMATION:
APPLICANT: PARANHOS-BACALA, Glaucia
APPLICANT: LESENECHAL, Mylene
APPLICANT: JOLIVET, Michel
TITLE OF INVENTION: NEW TRYPAOSOMA CRUZI ANTIGEN, AND GENE
TITLE OF INVENTION: ENCODING THE LATTEK; THEIR APPLICATION TO THE DETECTION OF
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Oliff & Berridge
STREET: 700 South Washington Street, Suite 300
CITY: Alexandria
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22314
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,917
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 36400
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3402 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-480-917-1

Query Match 100.0%; Score 976; DB 1; Length 3402;
Best Local Similarity 100.0%; Pred. No. 6; 9e-247;
Matches 976; Conservative 0; Mismatches 0; Gaps 0;
Oy 1 cagtgacgcgcgtacgcgttgcgttcacatcgtacgcgcgttagctgctcgtcgt 60

Db	1232	CAGGTACAGCGCTAACGGCGCTTTGCTTCATCTGACAGCGGAGCGGTAGCTGCTGAGCT	1291
Qy	61	gaactgtcgcattcgattgaaagcatcttccattccctgcgtctctcccgcaaggagaagaacagag	120
Db	1292	GACATGTGCGATTCCATTGAGACGATCTTCATCTCCGCTCTCCCGAGGGAAACAGACAG	1351
Qy	121	ccagagcaaaaaaatccgtcagtcagtcagcagcgcaagcggaagtcgtctccctcggcact	180
Db	1352	CCAGGCGCAAAAACATCGTAGTAGTGGGACGCGCAAACCGGGGTGTGTCTCTCGGGGCACT	1411
Qy	181	gaacgcgcgcagtagcagtcataccaaatacgaactctgcgcgtgtcgtacatccccctgcata	240
Db	1412	GACCGCGCGAGTATCACTCATATACCAATAGCACTTCTGCGCGTGTGCATCCCTCATCA	1471
Qy	241	ccccctcttcaagcgcaagcgcaagcgcgccctccctcgcgcgcgcgatctcgctag	300
Db	1472	CCCCCTGTTTACGGGCCAGCCAAAGGACGGCGGCGCTCTCCGCGCGCGGATCGGCTAG	1531
Qy	301	ccgcacgcctggggagcaaaatcatctgtcctaaactagtaatacagcttgaggatlaattacac	360
Db	1532	CCGCACGCGGGGAGACAGATCATTTGGTAAATCTAGTAATCACTGGGGATTAAATGTACAC	1591
Qy	361	caaaagcgcgctgcgtcagcactctggagcgccgcgcacgaagagttctacggcggtgcaccc	420
Db	1592	CAAAAGGAGCGCTGTGCACACTGTGAGCGCGCGCCACACAGAGGTCTATACGGCGGTGACGTC	1651
Qy	421	acgacataccgccccgcgcgcgcgaacaagtlcatalcagcgcgaatgycgcgcactgtgcgcgt	480
Db	1652	ACGACTATCCGCCCGCGACGCAACATTCATAGGGGCAACATGGCGCACTGTAGCGCT	1711
Qy	481	ggattgtgtgcagctaaatagtgctgcacagcgcgctcgtctcccaagccgcgcgcgaa	540
Db	1712	GGATTGTGTGCGAGCTATATAGTGTGTGCACGCGGCGCTCTCTCCACAGCGCGCGCAAA	1771
Qy	541	cccaacgaggaagaagaagcgctcgcgcgcgatitgaaacgagctcgtgtgcgataatgycg	600
Db	1772	CCAACACGAGAGAAAGAAAGCGCTCTCGCGGCAATGTGAACAGAGCTCCGTGGCGATAATACG	1831
Qy	601	acacgcgcgcgcgtctcaacacgcctctctccgcgcgcgcgcgcgaacgagatgycgtttgcgc	660
Db	1832	ACACGCCCGGGGCTTCACAAAGCGCTCTCTCCGACAGCGGCCCAAGCGGATGGGTTTGGCG	1891
Qy	661	gcagcagtatataccagtcgcgagggcgaggttcatcagtcgctgagcgagcttgaatcgcgt	720
Db	1892	GCACAGCTAATACAGTCGCGAGGGCGAGGTTCAATCACTGCTGAGACGGCTGGAGTGCCTC	1951
Qy	721	ataaccacaacgctcgcggtctgtgaagttgtctccctgaaacacatctcgaagagacatgaa	780
Db	1952	ATAACCAACACGCTCGCGGTCTGGAAGTGTCTCCCTGCACACCATTCGGAAGACATGAA	2011
Qy	781	caactcttgaaactggtgtttagaagcacagatgacagagctgcacgcagagcgcgtccaca	840
Db	2012	CAACTCTTGATCTGGGTTTAGAGCGACAGATGACAGAGACTCAGACAAGCGCTGCCAAC	2071
Qy	841	ccgcgaacacagcgcgcgagagcacaaagctctccgcgaaatacgcgtttgaagctaacac	900
Db	2072	CCGCAAAACAAGCCGAGAGACACAACTCCGGGAAATCATTCGTTTGAAGCGTCAACC	2131
Qy	901	ctgtgtctcatctgcgagatccctctctctgcacaatcatcaagaaagggttgaacgtgtgtgtg	960
Db	2132	CTGTGTCTCATTTGCGGATTCCTCTCTCCGCAACATCAGCAAGGGGCTGAAGGTGTGTG	2191
Qy	961	aacgagcgcatatgct 976	
Db	2192	AACGAGGCGCATTTATGT 2207	

```

1  APPLICANT: Nucleic acid molecules coding for mammalian
2  TITLE OF INVENTION: tumor suppressor proteins and methods for their isolation
3  NUMBER OF SEQUENCES: 15
4  COMPUTER READABLE FORM:
5  MEDIUM TYPE: Floppy disk
6  COMPUTER: IBM PC compatible
7  OPERATING SYSTEM: PC-DOS/MS-DOS
8  SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
9  CURRENT APPLICATION DATA:
10 APPLICATION NUMBER: US/08/718,661
11 FILING DATE:
12 CLASSIFICATION: 530
13 INFORMATION FOR SEQ ID NO: 1:
14 SEQUENCE CHARACTERISTICS:
15 LENGTH: 2790 base pairs
16 TYPE: nucleic acid
17 STRANDEDNESS: single
18 TOPOLOGY: linear
19 MOLECULE TYPE: cDNA to mRNA
20 FEATURE:
21 NAME/KEY: CDS
22 LOCATION: 542..2545
23 US-08-718-661-1

```

Query Match	4.0%	Score 39.2	DB 2	Length 2790
Best Local Similarity	54.1%	Pred. NO. 0.24		
Matches	80	Conservative	0	Mismatches 68
				Indels 0
				Gaps 0
QY 137	cggtagttgcagcagcgagaaaccgggggtgtgtgttcctcgggcacttgcagcgcgatagca	196		
Db 2114	CAGCAGCAGCAGCCAGCCGCACTGGCCGCGGCACACGACCGGTGCACCTGCACAGCGACTGGCA	2173		
QY 197	gtcataccaatacgaactcttcgcgctgtgcataccctctgcatacccccgttttagcgc	256		
Db 2174	GTGCCACAGCCACTTACCAACAGCCACAGATGCGACGACATTTCAGTTGAGATTCACAGCC	2233		
QY 257	cagccaaggcagcgcgctcctccgcgc	284		
Db 2234	CAGATGACAGTTCACACAGCTGCTGCGCG	2261		

RESULT 2
US-08-718-661-1
; Sequence 1, Application US/08718661
; Patent No. 5876972
; GENERAL INFORMATION:

RESULT 3
US-08-042-747A-7//C
Sequence 7, Application US/08042747A
Patent No. 5487969
GENERAL INFORMATION:
APPLICANT: Eberle, Richard
APPLICANT: Black, Darla
APPLICANT: Schinacriello, Franco
TITLE OF INVENTION: Cloning and Amplification of Monkey B
TITLE OF INVENTION: Cloning and Amplification of Monkey B
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cox & Smith Incorporated
STREET: 112 East Pecan Street, Suite 2000
CITY: San Antonio
STATE: Texas
COUNTRY: USA
ZIP: 78205
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/042,747A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Haymond, W. Bradley

SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/042,747A
 FILING DATE:
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Haymond, W. Bradley


```

: REGISTRATION NUMBER: 35186
: REFERENCE/DOCKET NUMBER: S-0072.179
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 210-554-5500
: TELEFAX: 210-226-8395
: TELEX: 767609
: INFORMATION FOR SEQ ID NO: 7:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2943 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: Genomic DNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 87..2744
: US-08-042-747A-7

Query Match          3.9%; Score 38.2; DB 1; Length 2943;
Best Local Similarity 47.7%; Pred. No. 0.45;
Matches 112; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

QY 209 cgactctcgcgtcgtcatccctcgtacacccctgtttcagcgccagccaagcagcag 268
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 255 CGTCCTCGTCCGGGACCTCTCGTCTCGGGAGCGCGGCGGTGGCGGCGGCGGCGG 196

QY 269 ccgcgcctctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 328
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 195 TCGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 136

QY 329 atctagtcatcagctgggggttaatgtcaccaaagagcgtctcagcactgtgagcgc 388
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 135 GGAGGACGGGGAGGGAAGAAGAGGGGGGGGTGCGCGAGCGCCGATCCCGGGGCGCG 76

QY 389 cgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 443
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 75 CGGCGCGTACCGCGGCGCGCTACGCGGCGGTGACATCCGCGCGCGCGCGCGGAGC 21

RESULT 4
US-08-474-379C-60
: Sequence 60, Application US/08474379C
: Patent No. 5977305
: GENERAL INFORMATION:
: APPLICANT: Wigler, Michael H.
: APPLICANT: Colicelli, John J.
: TITLE OF INVENTION: CLONING BY COMPLEMENTATION AND RELATED
: TITLE OF INVENTION: PROCESSES
: NUMBER OF SEQUENCES: 88
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
: STREET: 233 South Wacker Drive/6300 Sears Tower
: CITY: Chicago
: STATE: Illinois
: COUNTRY: United States of America
: ZIP: 60606-6402
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/474,379C
: FILING DATE: 07-JUN-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/511,715
: FILING DATE: 20-APR-1990
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/206,188
: FILING DATE: 01-MAR-1994
: PRIOR APPLICATION DATA:
```

```

: APPLICATION NUMBER: US 07/688,352
: FILING DATE: 19-APR-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: Clough, David W.
: REGISTRATION NUMBER: 36,107
: REFERENCE/DOCKET NUMBER: 27866/32771
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (312) 474-6300
: TELEFAX: (312) 474-0448
: INFORMATION FOR SEQ ID NO: 60:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2115 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 762..2115
: US-08-474-379C-60

Query Match          3.9%; Score 37.6; DB 2; Length 2115;
Best Local Similarity 63.0%; Pred. No. 0.59;
Matches 58; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 212 ctctcgcgcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 271
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 577 CTGCCCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 636

QY 272 ggcctcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 303
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 637 CCCCCCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 668

RESULT 5
US-09-146-249A-60
: Sequence 60, Application US/09146249A
: Patent No. 6069240
: GENERAL INFORMATION:
: APPLICANT: Wigler, Michael H.
: APPLICANT: Colicelli, John J.
: TITLE OF INVENTION: Cloning by Complementation and Related
: TITLE OF INVENTION: Processes
: NUMBER OF SEQUENCES: 85
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
: STREET: 6300 Sears Tower, 233 South Wacker Drive
: CITY: Chicago
: STATE: Illinois
: COUNTRY: United States of America
: ZIP: 60606-6402
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/146,249A
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/511,715
: FILING DATE: 20-APR-1990
: ATTORNEY/AGENT INFORMATION:
: NAME: Clough, David W.
: REGISTRATION NUMBER: 36,107
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 312/474-6300
: TELEFAX: 312-474-0448
: TELEX: 25-3856
: INFORMATION FOR SEQ ID NO: 60:
: SEQUENCE CHARACTERISTICS:
```



```

? ANTI-SENSE: NO
? ORIGINAL SOURCE:
? ORGANISM: Gossypium hirsutum
? STRAIN: Coker 312
? TISSUE TYPE: Fiber cells
? IMMEDIATE SOURCE:
? CLONE: H6
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 72..716
? FEATURE:
? NAME/KEY: sig_peptide
? LOCATION: 72..147
US-08-217-3

```

Query Match	3.88;	Score	36.6;	DB	1;	Length	913;
Best Local Similarity	58.98;	Pred.	No. 0.85;				
Matches	63;	Conservative	0;	Mismatches	44;	Indels	0;
						Gaps	0;

Qy 198 tcaataccaataagactcttcgcgctgtcatccctgatcaccaccttltacagccc 257
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 320 TCACCCAGTTCCTCCTCCGCAACTCCACCACCTGCCTTCCTCCTCGAACATCC 379

QY 258 agccaagycagccgcctcctgcgcgcgcgcatgcgtagccgc 304
 380 ACCTCAGCTTCTCCACCTCTGCACATCTCCACAGCTTCTCCAC 426
 Db

RESULT 11
US-07-885-970A-3

Sequence 3, Application US/07885970A
Patent No. 5495070
GENERAL INFORMATION:

TITLE OF INVENTION: GENETICALLY ENGINEERING COTTON
 NUMBER OF SEQUENCES: 33
 CORRESPONDENCE ADDRESS:

STREET: P.O. Box 2113, First Wisconsin Plaza
CITY: Madison
STATE: Wisconsin

ZIP: 53701
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk

```

;
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
;

```

FILING DATE: 19920518
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:

FILING DATE: 21-NOV-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/253,243

ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27,386

TELEPHONE: (608) 283-2478
TELEFAX: (608) 251-5139
INFORMATION FOR SEQ ID NO: 3:

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 913 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double

```

```

;      TOPOLOGY: linear
;      MOLECULE TYPE: CDNA
;      HYPOTHETICAL: NO
;

```

? ANTI-SENSE: NO
 ? ORIGINAL SOURCE:
 ? ORGANISM: *Gossypium hirsutum*
 ? STRAIN: Coker 312
 ? DEVELOPMENTAL STAGE: 15 day old fiber cells
 ? TISSUE TYPE: fiber cells
 ? IMMEDIATE SOURCE:
 ? LIBRARY: CKB15A1
 ? CLONE: H6
 ?
 ? US-07-885-970A-3

Query Match	3.8%	Score 36.6;	DB 1;	Length 913;
Best Local	Similarity 58.9%;	Pred. No. 0.85;		
Matches 63;	Conservative 0;	Mismatches 44;	Indels 0;	Gaps 0;

Qy 198 tcaataccaatacgaactcttcgcgctgtcgatccccctgatcacaccccttlttagcgcc 257
|| ||| || | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 320 TCACACCAGATTCTCCCTCCTCGCAACTGCACCACTGCTTCCTCCCTCCGCAACTCC 379

```

Oy      258 agccaagcgacgccgcgcctctctctgcccgcgcgcgcatcgctagaccgc 304
          | | | | | | | | | | | | | | | | | | | | |
Db      380 ACCTCAGCTTCTCCACCTCTGCACACTCTCCACCAACGACTTCTTCAC 426

```

RESULT 12
US-08-298-687A-3

```

; GENERAL INFORMATION:
;
; APPLICANT: John, Mallyakal E.

```

```

; TITLE OF INVENTION: GENETICALLY ENGINEERING COTTON
;
; TITLE OF INVENTION: PLANTS FOR ALTERED FIBER
;
; NUMBER OF SEQUENCES: 33

```

ADDRESS: Nicholas J. Seay, Quarles & Brady
STREET: P.O. Box 2113, First Wisconsin Plaza
CITY: Madison

COUNTRY: USA
ZIP: 53701
COMPUTER READABLE F

```
;
;
;      MODERN TITLE TOPPI CLASS
;
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Microsoft Word
```

CONTRACT NO. 687A
; APPLICATION NUMBER: US/08/298,687A
; FILING DATE:
; CLASSIFICATION: 800
;

APPLICATION NUMBER: US 07/617,239
FILING DATE: 21-NOV-1990
PRIOR APPLICATION DATA:

FILED DATE: 04-OCT-1988
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.

REGISTRATION NUMBER: 27,000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 283-2478
TELEFAX: (608) 251-5139

```

; INFORMATION FOR SEQ ID NO: 3:
;
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 913 base pairs
;   TYPE: nucleic acid
;

```

```

; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO

```

```

; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Gossypium hirsutum
;

```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 21, 2001, 02:15:37 : Search time 6407.28 Seconds
(without alignments)
2250.903 Million cell updates/sec

Title: US-09-138-735-1_COPY_1232_2207
Perfect score: 976
Sequence: 1 caggtacagcgtacgctt.....tltgacagcgccattatgt 976

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 14155048 seqs, 7388405095 residues

Total number of hits satisfying chosen parameters: 28310096

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Pending_Patents_NA_Main:*

1: /cgnl_7/ptodata/1/pna/PCYUS.COMB.seq:*
2: /cgnl_7/ptodata/1/pna/US06.COMB.seq:*
3: /cgnl_7/ptodata/1/pna/US07.COMB.seq:*
4: /cgnl_7/ptodata/1/pna/US08.COMB.seq:*
5: /cgnl_7/ptodata/1/pna/US081.COMB.seq:*
6: /cgnl_7/ptodata/1/pna/US082.COMB.seq:*
7: /cgnl_7/ptodata/1/pna/US083.COMB.seq:*
8: /cgnl_7/ptodata/1/pna/US084.COMB.seq:*
9: /cgnl_7/ptodata/1/pna/US085.COMB.seq:*
10: /cgnl_7/ptodata/1/pna/US086.COMB.seq:*
11: /cgnl_7/ptodata/1/pna/US087.COMB.seq:*
12: /cgnl_7/ptodata/1/pna/US088.COMB.seq:*
13: /cgnl_7/ptodata/1/pna/US089.COMB.seq:*
14: /cgnl_7/ptodata/1/pna/US090.COMB.seq:*
15: /cgnl_7/ptodata/1/pna/US091.COMB.seq:*
16: /cgnl_7/ptodata/1/pna/US092.COMB.seq:*
17: /cgnl_7/ptodata/1/pna/US093.COMB.seq:*
18: /cgnl_7/ptodata/1/pna/US094.COMB.seq:*
19: /cgnl_7/ptodata/1/pna/US095A.COMB.seq:*
20: /cgnl_7/ptodata/1/pna/US095B.COMB.seq:*
21: /cgnl_7/ptodata/1/pna/US095C.COMB.seq:*
22: /cgnl_7/ptodata/1/pna/US095D.COMB.seq:*
23: /cgnl_7/ptodata/1/pna/US096A.COMB.seq:*
24: /cgnl_7/ptodata/1/pna/US096B.COMB.seq:*
25: /cgnl_7/ptodata/1/pna/US096C.COMB.seq:*
26: /cgnl_7/ptodata/1/pna/US096D.COMB.seq:*
27: /cgnl_7/ptodata/1/pna/US096E.COMB.seq:*
28: /cgnl_7/ptodata/1/pna/US097A.COMB.seq:*
29: /cgnl_7/ptodata/1/pna/US097B.COMB.seq:*
30: /cgnl_7/ptodata/1/pna/US097C.COMB.seq:*
31: /cgnl_7/ptodata/1/pna/US098.COMB.seq:*
32: /cgnl_7/ptodata/1/pna/US099.COMB.seq:*
33: /cgnl_7/ptodata/1/pna/US6000.COMB.seq:*
34: /cgnl_7/ptodata/1/pna/US6001.COMB.seq:*
35: /cgnl_7/ptodata/1/pna/US6002.COMB.seq:*
36: /cgnl_7/ptodata/1/pna/US6003.COMB.seq:*
37: /cgnl_7/ptodata/1/pna/US6004.COMB.seq:*
38: /cgnl_7/ptodata/1/pna/US6005.COMB.seq:*
39: /cgnl_7/ptodata/1/pna/US6006.COMB.seq:*
40: /cgnl_7/ptodata/1/pna/US6007.COMB.seq:*
41: /cgnl_7/ptodata/1/pna/US6008.COMB.seq:*
42: /cgnl_7/ptodata/1/pna/US6009.COMB.seq:*
43: /cgnl_7/ptodata/1/pna/US6011.COMB.seq:*

44: /cgnl_7/ptodata/1/pna/US6012.COMB.seq:*
45: /cgnl_7/ptodata/1/pna/US6013.COMB.seq:*
46: /cgnl_7/ptodata/1/pna/US6014.COMB.seq:*
47: /cgnl_7/ptodata/1/pna/US6015.COMB.seq:*
48: /cgnl_7/ptodata/1/pna/US6016.COMB.seq:*
49: /cgnl_7/ptodata/1/pna/US6017.COMB.seq:*
50: /cgnl_7/ptodata/1/pna/US6018.COMB.seq:*
51: /cgnl_7/ptodata/1/pna/US6019.COMB.seq:*
52: /cgnl_7/ptodata/1/pna/US6020.COMB.seq:*
53: /cgnl_7/ptodata/1/pna/US6021.COMB.seq:*
54: /cgnl_7/ptodata/1/pna/US6022.COMB.seq:*
55: /cgnl_7/ptodata/1/pna/US6023.COMB.seq:*
56: /cgnl_7/ptodata/1/pna/US6024.COMB.seq:*
57: /cgnl_7/ptodata/1/pna/US6025.COMB.seq:*
58: /cgnl_7/ptodata/1/pna/US6026.COMB.seq:*
59: /cgnl_7/ptodata/1/pna/US6027.COMB.seq:*
60: /cgnl_7/ptodata/1/pna/US6028.COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	976	100.0	3402	13 US-08-988-242-1	Sequence 1, Appl1
2	976	100.0	3402	15 US-09-138-736-1	Sequence 1, Appl1
3	46	4.7	606	27 US-09-699-999-5885	Sequence 5885, Ap
4	45.8	4.7	470	27 US-09-699-999-4102	Sequence 4102, Ap
5	44.6	4.6	599	17 US-09-396-087-4576	Sequence 4576, Ap
6	42.6	4.4	6257	24 US-09-620-392-44819	Sequence 44819, A
7	42.6	4.4	29301	28 US-09-702-134-5721	Sequence 5721, Ap
8	42.4	4.3	600	52 US-60-209-830-45296	Sequence 45296, A
9	42.2	4.3	491	51 US-60-197-872-57273	Sequence 57273, A
10	42.2	4.3	8953	24 US-09-620-392-68153	Sequence 68153, A
11	42.2	4.3	10344	24 US-09-620-392-65686	Sequence 65686, A
12	42.2	4.3	10348	28 US-09-702-134-8344	Sequence 8344, Ap
13	42	4.3	7227	48 US-60-168-139-2273	Sequence 2273, Ap
14	42	4.3	26276	19 US-09-514-000-337	Sequence 337, App
15	42	4.3	397803	29 US-09-739-449-217	Sequence 217, App
16	41.6	4.3	600	52 US-60-207-458-67549	Sequence 67549, A
17	41.4	4.2	322	28 US-09-716-472-1859	Sequence 1859, Ap
18	41.2	4.2	728	23 US-09-614-387-7832	Sequence 7832, Ap
19	41.2	4.2	728	28 US-09-716-472-5422	Sequence 5422, App
20	41.2	4.2	741	50 US-60-185-213-686	Sequence 686, App
21	41.2	4.2	12871	24 US-09-620-392-3460	Sequence 3460, App
22	41.2	4.2	33579	28 US-09-702-134-4946	Sequence 4946, Ap
23	41.2	4.2	399	52 US-60-207-458-74910	Sequence 74910, A
24	40.8	4.2	415	52 US-60-207-458-150303	Sequence 150303, A
25	40.8	4.2	547	52 US-60-207-458-149725	Sequence 149725, A
26	40.8	4.2	573	22 US-09-553-094-5427	Sequence 5427, Ap
27	40.8	4.2	635	52 US-60-209-830-60545	Sequence 60545, A
28	40.8	4.2	637	52 US-60-207-458-151348	Sequence 151348, A
29	40.8	4.2	668	52 US-60-207-458-151301	Sequence 151301, A
30	40.8	4.2	550	51 US-60-197-872-44492	Sequence 44492, A
31	40.6	4.2	470	25 US-09-654-617-111007	Sequence 111007, A
32	40.4	4.1	470	25 US-09-684-016-111007	Sequence 111007, A
33	40.4	4.1	477	15 US-09-135-994-1	Sequence 1, Appl1
34	40.4	4.1	569	53 PCT-US01-750-30151	Sequence 30151, A
35	40.4	4.1	5537	18 PCT-US01-040988-2481	Sequence 2481, Ap
36	40.4	4.1	5537	18 US-09-496-914A-6707	Sequence 6707, Ap
37	40.4	4.1	5537	22 US-09-560-875A-6707	Sequence 6707, Ap
38	40.4	4.1	6248	1 PCT-US01-040988-513	Sequence 513, App
39	40.4	4.1	4403765	15 US-09-103-840A-2	Sequence 2, Appl1
40	40.4	4.1	441529	15 US-09-103-840A-2	Sequence 392521, A
41	40.4	4.1	811	25 US-09-654-617-392521	Sequence 392521, A
42	40.2	4.1	811	27 US-09-684-016-392521	Sequence 392521, A
43	40.2	4.1	6867	24 US-09-620-392-66151	Sequence 66151, A
44	40	4.1	7443	24 US-09-620-392-55872	Sequence 55872, A
45	40	4.1			

ALIGNMENTS

RESULT 1

US-08-988-242-1

; Sequence 1, Application US/08988242

; GENERAL INFORMATION:

; APPLICANT: PARANHOS-BACCALA, GLAUCIA

; APPLICANT: LESENCHAL, MYLENE

; APPLICANT: JOLIVET, MICHEL

; APPLICANT: MANDRAND, BERNARD

; TITLE OF INVENTION: NEW TRYPANOSOMA CRUZI ANTIGEN, GENE

; TITLE OF INVENTION: ENCODING THEREFOR, AND METHODS OF DETECTING AND TREATING

; NUMBER OF SEQUENCES: 19

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: OLITF & BERRIDGE, PLC

; STREET: P.O. BOX 19928

; CITY: Alexandria

; STATE: Virginia

; COUNTRY: U.S.A.

; ZIP: 22320

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/988,242

; FILING DATE: 10-DEC-1997

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Berridge, William P.

; REGISTRATION NUMBER: 30,024

; REFERENCE/DOCKET NUMBER: WPP 36400A

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 703-836-6400

; TELEFAX: 703-836-2787

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 3402 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: CDNA

US-08-988-242-1

Query Match 100.0%; Score 976; DB 13; Length 3402;

Best Local Similarity 100.0%; Pred. No. 2,8e-221;

Matches 976; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caaggtacaggtatgagcttctgctcaatcgtacacgcaaggtgctctcctgagct 60
DB 1232 CAGGTACAGGTATGAGCTTCTGCTCAATCTGACACGCCAGGTACTGCTGCTGCT 1291
QY 61 gacatgcatcgtatgacgacatcctcctcgtcctcccgcaaggaagaacagcag 120
DB 1292 GACATGTCATTCGATTGACGATCTTCATCTCCGGTCTCCGCAAGGAAGACAGCAG 1351
QY 121 cgaagcacaacacatcgtatgctgcaagcggaacggggtgtgtgtcctcgggagc 180
DB 1352 CCAAGCCCAACCAATCGTATGTCGCGACGCGCAACCGGGGTGTGTCTCTCGGCGACT 1411
QY 181 gacggcgaggtgaggtatcaccatgactctgctcgtctgctatccctctgacat 240
DB 1412 GACGCGCGGAGTACGATCTACCAATACGACTTCTGCTCTGCTGATCCCTGCACTCA 1471
QY 241 cccctgttctcagcgcaagcgcaagcgagcgagcctcctcagcgcgagcagctgag 300
DB 1472 CCCCCTTTTCAAGCGCAGCAGCAGCGCGCGCTCTCTCCCGCGGCGCATCGCTGAG 1531

QY 301 ccgacgtggtggagcaagatcgtactatctagtcaatcagctggggaatgacac 360
DB 1532 CCGACGTGGGAGCAAGATCATTTGCTAATCTAGTAATCAGCTGGGATTAATGTCACC 1591
QY 361 caaagagcgctgctcagcacttgagcgccggcgcaagcaggtctcagcggtgacgtcc 420
DB 1592 CAAAGAGCGCTCTCAGCACTGAGCGCGCGCCACACAGAGGTCTAGCGGTGAGCTCC 1651
QY 421 acgactacgcccccgagcgaaacagatccatcagggcacaatggtcgactgtgagcgtc 480
DB 1652 ACGACTACCGCCCCGACGCAACCAATCCATACGGGACAAATGGCCGACTGTGAGCGCT 1711
QY 481 ggtgtgtgcaagctaatagtgtgctcagcgcgccctgctcccaagcgcgagaa 540
DB 1712 GGATGTGTGACACTAATAGTGTGTCACAGCGCGCTCTCTCCACACCGCGGAGAA 1771
QY 541 ccaacagaggaagaaagcctcctcgcgcatgtgaaagcagctccgctgagcgaatgtgc 600
DB 1772 CCAACAGAGAGAAAGAGCGCTCCGCGCATGTGAACAGAGCTCCGTGCGATAAATGCG 1831
QY 601 acagcccgagccttcacacagcctctcctcccgagcgcaagagatgagcttggcg 660
DB 1832 ACAGCGCGCGCGCTTACACAGCGCTCTCTCCCGACGCGCAACGAGATGCGCTTGGCG 1891
QY 661 gcagcagatatacagtcgagggcgaggtcatalcagctgagcgagctgagctgc 720
DB 1892 GCAGCGATATACAGTCGAGCGGCGAGGTTCATCATCTGAGCGCGCTGAGTCCGTC 1951
QY 721 ataccacaagcgtctgaggtctgaaagttgctccctgaacacatctgaagagacatgaa 780
DB 1952 ATACCAACACGTCGTGGGTTGAAAGTTCCTCTGACACCAATGGAAGACCAATGAA 2011
QY 781 caactctgatatctggtgtttagagcacagatgacagagctgcaagagcgctcccaaa 840
DB 2012 CAACTTCTGAATCTGGGTTAGAGGACACAGATGACAGAGTGCAGAGCGCTTCAACA 2071
QY 841 cgcgaacacacgcgagagagacacacagctccggaatcatcgtgtttgagcgtaccc 900
DB 2072 CCGCAACACACGCGAGAGACACACACCTCCGCAAAATCATCTCTTTAGAGAGTACACC 2131
QY 901 ctgtgtctatctgagatctcctctcgtcgaacatcagcaaggggtgagcgtggtg 960
DB 2132 CTGTGTCTCTCTCGGATTCCTCTCTCTCGCAACATCAGCAAGGGGTGAAGCGTGTG 2191
QY 961 aacgagccatattgt 976
DB 2192 AACGAGCCATTATGT 2207

RESULT 2

US-09-138-736-1

; Sequence 1, Application US/09138736

; GENERAL INFORMATION:

; APPLICANT: PARANHOS-BACCALA, GLAUCIA

; APPLICANT: LESENCHAL, Mylene

; TITLE OF INVENTION: NEW TRYPANOSOMA CRUZI ANTIGEN, AND GENE

; TITLE OF INVENTION: ENCODING THE LATTER; THEIR APPLICATION TO THE DETECTION OF

; NUMBER OF SEQUENCES: 13

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Olitf & Berridge

; STREET: 700 South Washington Street, Suite 300

; CITY: Alexandria

; STATE: Virginia

; COUNTRY: U.S.A.

; ZIP: 22314

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

[illegible]

```

RESULTS      7
US-09-702-134-5721
: Sequence 5721, Application US/09702134
: GENERAL INFORMATION:
: APPLICANT: Boukharov, Andrey A.
: APPLICANT: Cao, Tongwei
: APPLICANT: Kovalic, David K.
: APPLICANT: Liu, Jingtong
: APPLICANT: McIninch, James
: APPLICANT: Wu, Wei
: TITLE OF INVENTION: Plant Genome Sequence and Uses Thereof
: FILE REFERENCE: 38-21(51237)F
: CURRENT APPLICATION NUMBER: US/09/702,134
: CURRENT FILING DATE: 2000-10-31
: NUMBER OF SEQ ID NOS: 52202
: SEQ ID NO 5721
: LENGTH: 29501
: TYPE: DNA
: ORGANISM: Oryza sativa
US-09-702-134-5721

```

Query Match	4.4%	Score 42.6	DB 28	Length 29301
Best Local Similarity	48.2%	Pred. No. 12		
Matches 120	Conservative 0	Mismatches 129	Indels 0	Gaps 0

[illegible]

RESULT 8
US-60-209-830-45296
; Sequence 45296, Application US/60209830
; GENERAL INFORMATION:

```

: APPLICANT: Andersen, Scott E.
: APPLICANT: Castiglioni, Paolo
: APPLICANT: Hardeman, Kristine J.
: APPLICANT: Kerk, Nancy M.
: APPLICANT: La Rosa, Thomas J.
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
: TITLE OF INVENTION: PLANTS
: FILE REFERENCE: 38-21(51934)A
: CURRENT APPLICATION NUMBER: US/60/209,830
: CURRENT FILING DATE: 2000-06-06
: NUMBER OF SEQ ID NOS: 62628
: SEQ ID NO 45296
: LENGTH: 600
: TYPE: DNA
: ORGANISM: Zea mays
: OTHER INFORMATION: Clone ID: LIB3596-038-Q6-K6-G7
US-60-209-830-45296

```

[illegible]

```

RESULT          9
US-60-197-872-57273/c
: Sequence: 57273, Application US/60197872
: GENERAL INFORMATION:
: APPLICANT: Bougri, Olegs
: APPLICANT: Byrum, Joseph R.
: APPLICANT: De la Pena, Robert C.
: APPLICANT: La Rosa, Thomas J.
: APPLICANT: Shukla, Hridayabhiranjan
: TITLE OF INVENTION: Nucleic acid molecules and other molecules associated w
: TITLE OF INVENTION: plants
: FILE REFERENCE: 38-21(51892)A
: CURRENT APPLICATION NUMBER: US/60/197,872
: CURRENT FILING DATE: 2000-04-19
: NUMBER OF SEQ ID NOS: 76255
: SEQ ID NO 57273
: LENGTH: 491
: TYPE: DNA
: ORGANISM: Oryza sativa nipponbare
: OTHER INFORMATION: Clone ID: jC-osfLIB3479034a01b1
US-60-197-872-57273

```

	Query Match	Similarity	4.3%	Score 42.2	DB 51	Length 491
	Best Local	Similarity	53.3%	Pred. No. 8.4	Mismatches 78	Indels 0
	Matches	89	Conservative	0	Mismatches	Gaps
Qy	142	gtggagacgcgcaaacacggggtgtgtgtcctcgagcaactgacgcggcgatagcagtcac	201			
Db	371	gCGCGAGGGTACAGCGCGGCTGTGTGGTTCCGAGGTAAGGAGAAACGGTTCCAGTCCC	312			
Qy	202	accaatagcaactctgtccgctgtctgtacatccctgtacataccccctgtttcaagccacgac	261			
Db	311	GCCTCTTCGCGACACTCGTACCTGCTCAGCCCGCGCGCGCCGACATGCTGCCACGG	252			
Qy	262	aaggaagcgcgcgcctcctctgcgcgcgcgcgatctggtgagccgacgt	308			
Db	251	CGTGGGCGCTGGCTCTTGTGCACCTCTGCTGCGCTCGCACATCGAGGT	205			

RESULT 10

```
US-09-620-392-68153/c
; Sequence 68153, Application US/09620392
; GENERAL INFORMATION:
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: McIninch, James
; TITLE OF INVENTION: Plant Genome Sequence And Uses Thereof
; FILE REFERENCE: 38-21(51237)E
; CURRENT APPLICATION NUMBER: US/09/620,392
; CURRENT FILING DATE: 2000-07-19
; NUMBER OF SEQ ID NOS: 69652
; SEQ ID NO 68153
; LENGTH: 8953
; TYPE: DNA
; ORGANISM: Oryza sativa
US-09-620-392-68153

Query Match
Best Local Similarity 4.3%; Score 42.2; DB 24; Length 8953;
Matches 95; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 133 acatcgtagtggcagcgagaaacggggtgtgtctctcgggcaactgacggcgaggt 192
DB 1548 ACACCGCGCCTCGCTCCAGCACCGTCCGCTTCGGCCGCGGCATCCCGCGCGCGGT 1489
QY 193 agcagtcatacaatacgaactctgcccgtctgcatcccttgatcaaccctgtttca 252
DB 1488 ACAGTCCGACACCGCGGGTTCGGCTCACCCCTCTCTCAGACACCGCACCCCGCCT 1429
QY 253 gcgcagcagcaagcgagcgccctctctgcccggcgcgatcgatcgacgcagctggg 312
DB 1428 CCTCCATCCACGACGCTCCGCCGCCGCCGCCGACACCGCACCCCGCGCGGCC 1369
QY 313 agc 315
DB 1368 ATC 1366

RESULT 11
US-09-620-392-65686
; Sequence 65686, Application US/09620392
; GENERAL INFORMATION:
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: McIninch, James
; TITLE OF INVENTION: Plant Genome Sequence And Uses Thereof
; FILE REFERENCE: 38-21(51237)E
; CURRENT APPLICATION NUMBER: US/09/620,392
; CURRENT FILING DATE: 2000-07-19
; NUMBER OF SEQ ID NOS: 69652
; SEQ ID NO 65686
; LENGTH: 10344
; TYPE: DNA
; ORGANISM: Oryza sativa
US-09-620-392-65686

Query Match
Best Local Similarity 4.3%; Score 42.2; DB 24; Length 10344;
Matches 95; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 133 acatcgtagtggcagcgagaaacggggtgtgtctctcgggcaactgacggcgaggt 192
DB 8556 acacccgctgctgctcagacacgctgctgctgctgctgctgctgctgctgctgctgct 8615
QY 193 agcagtcatacaatacgaactctgcccgtctgcatcccttgatcaaccctgtttca 252
DB 8616 acgagtcgacacgcgcgggtctcgcctacccctctctcagacacgcaccccgccct 8675
QY 253 gcgcagcagcaagcgagcgccctctctgcccggcgcgatcgatcgacgcagctggg 312
```

```
DB 8676 cctccatccagacgctccgcccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 8735
QY 313 agc 315
DB 8736 atc 8738

RESULT 12
US-09-702-134-8344/c
; Sequence 8344, Application US/09702134
; GENERAL INFORMATION:
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Cao, Yongwei
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: McIninch, James
; APPLICANT: Wu, Wei
; TITLE OF INVENTION: Plant Genome Sequence and Uses Thereof
; FILE REFERENCE: 38-21(51237)F
; CURRENT APPLICATION NUMBER: US/09/702,134
; CURRENT FILING DATE: 2000-10-31
; NUMBER OF SEQ ID NOS: 52202
; SEQ ID NO 8344
; LENGTH: 10348
; TYPE: DNA
; ORGANISM: Oryza sativa
US-09-702-134-8344

Query Match
Best Local Similarity 4.3%; Score 42.2; DB 28; Length 10348;
Matches 95; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 133 acatcgtagtggcagcgagaaacggggtgtgtctctcgggcaactgacggcgaggt 192
DB 1793 ACACCCGCGCTCGCTCCAGCACCGTCCGCTTCGGCCGCGGCATCCCGCGCGCGGT 1734
QY 193 agcagtcatacaatacgaactctgcccgtctgcatcccttgatcaaccctgtttca 252
DB 1733 ACAGTCCGACACCGCGGGTTCGGCTCACCCCTCTCTCAGACACCGCACCCCGCCT 1674
QY 253 gcgcagcagcaagcgagcgccctctctgcccggcgcgatcgatcgacgcagctggg 312
DB 1673 CCTCCATCCACGACGCTCCGCCGCCGCCGCCGACACCGCACCCCGCGCGGCC 1614
QY 313 agc 315
DB 1613 ATC 1611

RESULT 13
US-60-168-139-2273
; Sequence 2273, Application US/60168139
; GENERAL INFORMATION:
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequence And Uses Thereof
; FILE REFERENCE: 38-21(15490)A
; CURRENT APPLICATION NUMBER: US/60/168,139
; CURRENT FILING DATE: 1999-12-01
; NUMBER OF SEQ ID NOS: 3432
; SEQ ID NO 2273
; LENGTH: 7227
; TYPE: DNA
; ORGANISM: Agrobacterium tumefaciens
; FEATURE:
; OTHER INFORMATION: unsure at all n locations
US-60-168-139-2273

Query Match
Best Local Similarity 4.3%; Score 42; DB 48; Length 7227;
Matches 95; Conservative 0; Mismatches 88; Indels 0; Gaps 0;
```

[illegible]

```

RESULT 14
US-09-514-000-337/c
: Sequence 337, Application US/09514000
: GENERAL INFORMATION:
: APPLICANT: Hinkle, Gregory J.
: APPLICANT: Slater, Steven C.
: TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
: FILE REFERENCE: 38-10(115490)B
: CURRENT APPLICATION NUMBER: US/09/514,000
: CURRENT FILING DATE: 2000-02-23
: NUMBER OF SEQ ID NOS: 15034
: SEQ ID NO 337
: LENGTH: 26276
: TYPE: DNA
: ORGANISM: Agrobacterium tumefaciens
US-09-514-000-337

```

Query Match	4.3%	Score 42	DB 19	Length 26276
Best Local Similarity	60.5%	Pred. No. 16		
Matches 69	Conservative 0	Mismatches 45	Indels 0	Gaps 0

[illegible]

```

RESULT 15
US-09-739-449-217
: Sequence 217, Application US/09739449
: GENERAL INFORMATION:
: APPLICANT: Hinkle, Gregory J.
: APPLICANT: Slater, Steven C.
: TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
: FILE REFERENCE: 38-10(15490)C
: CURRENT APPLICATION NUMBER: US/09/739,449
: CURRENT FILING DATE: 2000-12-19
: PRIOR APPLICATION NUMBER: US 09/514,000
: PRIOR FILING DATE: 2000-02-23
: NUMBER OF SEQ ID NOS: 13351
: SEQ ID NO 217
: LENGTH: 397803
: TYPE: DNA
: ORGANISM: Agrobacterium tumefaciens
: FEATURE:
: NAME/KEY: unsure
: LOCATION: (1)..(397803)
: OTHER INFORMATION: unsure at all n locations
US-09-739-449-217

```

	Query Match	4.3%	Score 42	DB 29	Length 397803;
	Best Local Similarity	60.5%;	Pred. NO. 23;		
	Matches	69; Conservative	0; Mismatches	45; Indels	0; Gaps
Dy	183	cgcggcagatgacgatcaccaatagaactcttcgcgtctcatccacctgacacc	242		
Db	66662	ccccccaattcglttcccacaacttgacacctccgcgcgtagcgtagcagccctccagacc	66741		

QY 243 ccctgtttcagcgccacgccaagggcagccgcctctctcgcgcgcgatgcgc 296
 ||| | ||| ||| | ||| ||||| ||||| | | ||
Db 66742 ggaagtgcgtcacgcgcgtagtattgtctctccagccgcgcgtccaccgcgc 66795

```
Search completed: September 21, 2001, 02:19:42
Job time: 30454 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 21, 2001, 02:35:38 : Search time 1035.34 Seconds
(without alignments)
1881.359 Million cell updates/sec

Title: US-09-138-735-1_COPY_1232_2207
Perfect score: 976
Sequence: 1 caggtacagcgtaacgctt.....tgytaacgagccatattgt 976

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1741792 seqs, 997871973 residues

Total number of hits satisfying chosen parameters: 3483584

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_New:*
1: /cgnl_7/ptodata/2/pna/PCW_NEW_COMB.seq:*
2: /cgnl_7/ptodata/2/pna/US06_NEW_COMB.seq:*
3: /cgnl_7/ptodata/2/pna/US07_NEW_COMB.seq:*
4: /cgnl_7/ptodata/2/pna/US08_NEW_COMB.seq:*
5: /cgnl_7/ptodata/2/pna/US09_NEW_COMB.seq:*
6: /cgnl_7/ptodata/2/pna/US09_NEW_COMB.seq1:*
7: /cgnl_7/ptodata/2/pna/US09_NEW_COMB.seq2:*
8: /cgnl_7/ptodata/2/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	976	100.0	3402	6	US-09-138-735-1
2	42	4.3	397803	6	US-09-803-110-217
3	40.4	4.1	477	7	US-09-684-843A-1
4	38.8	4.0	657	7	US-09-922-460-3
5	38.8	4.0	657	7	US-09-923-236-3
6	38.8	4.0	657	7	US-09-922-469-3
7	38.6	4.0	411	7	US-09-902-540-5862
8	38.6	4.0	1209	7	US-09-902-540-147
9	38.2	3.9	495	8	US-60-253-651-24827
10	38.2	3.9	670	8	US-60-253-651-24775
11	37.8	3.9	405	8	US-60-253-654-1232
12	37.8	3.9	405	8	US-60-253-654-1232
13	37.8	3.9	578	8	US-60-253-654-18403
14	37.8	3.9	578	8	US-60-253-654-18403
15	37.8	3.9	608	8	US-60-253-654-18342
16	37.8	3.9	608	8	US-60-253-654-18342
17	37.4	3.8	1039	8	US-60-278-561-9647
18	37.4	3.8	6712	8	US-60-278-232-2818
19	37.2	3.8	522	6	US-09-823-241-9418
20	37.2	3.8	2406	6	US-09-760-466-345
21	37.2	3.8	6483	7	US-09-880-107-3699
22	37.2	3.8	10733	7	US-09-902-540-1021
23	36.8	3.8	435	5	US-09-842-827-1975
24	36.8	3.8	435	7	US-09-912-292-40706
25	36.8	3.8	621	7	US-09-840-424-56

26	36.6	3.8	30783	7	US-09-902-540-1258	Sequence 1258, Ap
27	36.4	3.7	819	7	US-09-902-540-8196	Sequence 8196, Ap
28	36.4	3.7	1020	1	PCR-US01-08631-12729	Sequence 12729, A
29	36.4	3.7	7190	7	US-09-902-540-843	Sequence 843, App
30	36.2	3.7	389	6	US-09-652-125A-2957	Sequence 2957, Ap
31	36.2	3.7	692	6	US-09-866-555-15847	Sequence 15847, A
32	36.2	3.7	2405	1	PCR-US01-08631-9988	Sequence 9988, Ap
33	36.2	3.7	334	7	US-09-850-147-15625	Sequence 15625, Ap
34	36	3.7	483	6	US-09-795-300-42	Sequence 42, Appl
35	36	3.7	488	7	US-09-834-366-2649	Sequence 2649, A
36	36	3.7	505	5	US-09-846-039-59	Sequence 59, Appl
37	36	3.7	592	8	US-60-253-654-24123	Sequence 24123, A
38	36	3.7	592	8	US-60-253-654-24123	Sequence 24123, A
39	36	3.7	1448	5	US-09-925-298-318	Sequence 318, App
40	36	3.7	16080	7	US-09-724-568-48	Sequence 48, Appl
41	36	3.7	16080	7	US-09-724-568-48	Sequence 48, Appl
42	36	3.7	16080	7	US-09-724-571-48	Sequence 48, Appl
43	35.8	3.7	446	7	US-09-912-292-19829	Sequence 19829, A
44	35.6	3.6	299	7	US-09-859-490-2340	Sequence 2340, Ap
45	35.6	3.6	456	6	US-09-904-809-19195	Sequence 19195, A

ALIGNMENTS

RESULT 1
US-09-138-735-1
Sequence 1, Application US/09138735
GENERAL INFORMATION:
APPLICANT: PARANHOS-BACCALA, Glaucia
APPLICANT: LEBENECHAL, Mylene
APPLICANT: JOLIVET, Michel
TITLE OF INVENTION: TRYPANOSOMA CRUZI ANTIGEN. GENE ENCODING THEREFOR AND METHODS
FILE REFERENCE: WPB 36400B
CURRENT FILING DATE: US/09/138/735
CURRENT FILING DATE: 1998-08-24
PRIOR APPLICATION NUMBER: US 08/480, 917
PRIOR FILING DATE: 1995-06-07
PRIOR APPLICATION NUMBER: FR 94/10132
PRIOR FILING DATE: 1994-08-12
NUMBER OF SEQ ID NOS: 13
SOFTWARE: PatentIn version 3.0
SEQ ID NO 1
LENGTH: 3402
TYPE: DNA
ORGANISM: Trypanosoma cruzi
US-09-138-735-1

Query Match 100.0%, Score 976, DB 6, Length 3402;
Best Local Similarity 100.0%, Pred. No. 1.6e-228;
Matches 976; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 caggtacagcgtaacgcttcttcacatcgtaacgagcgagcgtagctgctcgtgct 60
1232 caggtacagcgtaacgcttcttcacatcgtaacgagcgagcgtagctgctcgtgct 1291
61 gacatgtcatttcgatttcgacatcttcacatcttcgcttcctcccgaggaagacagcag 120
1292 gacatgtcatttcgatttcgacatcttcacatcttcgcttcctcccgaggaagacagcag 1351
121 ccaggtacagcgtaacgcttcttcacatcgtaacgagcgagcgtagctgctcgtgct 180
1352 ccaggtacagcgtaacgcttcttcacatcgtaacgagcgagcgtagctgctcgtgct 1411
181 gacgtgagcgtagcgatcattacacatcgtaacgagcgtagctgctcgtgctcgtgct 240
1412 gacgtgagcgtagcgatcattacacatcgtaacgagcgtagctgctcgtgctcgtgct 1471
241 cccctgttccagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 300
1472 cccctgttccagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 1531

QY	301	ccgcacgcggggagagaagatcatgtgtctaattagtgaatcagctctggggatcaatgaccc	360
Db	1532	ccgcacgcggggagagaagatcatgtgtctaattagtgaatcagctctggggatcaatgaccc	1591
QY	361	caaaagagagctctctcagcactctgagagccgcggccacagcagagctctacgctgctgacgtcc	420
Db	1592	caaaagagagctctcagcactctgagagccgcggccacagcagagctctacgctgctgacgtcc	1651
QY	421	acgactaccgcccgcagcagcaaaagctctccatacagggacaaatgctcgagctctgtacgct	480
Db	1652	acgactaccgcccgcagcagcaaaagctctccatacagggacaaatgctcgagctctgtacgct	1711
QY	481	ggatctggtgagcagctaaatagctgtgcacgacgagctctgtctccacacagccgcgcgaaa	540
Db	1712	ggatctggtgagcagctaaatagctgtgcacgacgagctctgtctccacacagccgcgcgaaa	1771
QY	541	ccaacagagagaagaaagagctctcgaggcatgtgaacagagctccgtggtcgataaagtcg	600
Db	1772	ccaacagagagaagaaagagctctcgaggcatgtgaacagagctccgtggtcgataaagtcg	1831
QY	601	acaagccggcgccctccacaacagccctctcccgacgagcgcaacagagatggtgtttgctg	660
Db	1832	acaagccggcgccctccacaacagccctctcccgacgagcgcaacagagatggtgtttgctg	1891
QY	661	gcagcagatataccagctcggagggcgaagttcatcatgctctgtgagcggcttgagctcgtc	720
Db	1892	gcagcagatataccagctcggagggcgaagttcatcatgctctgtgagcggcttgagctcgtc	1951
QY	721	ataacaaacagctctcgaggtctctgaagtgtctccctgacacacatctcgaaagagacatgaa	780
Db	1952	ataacaaacagctctcgaggtctctgaagtgtctccctgacacacatctcgaaagagacatgaa	2011
QY	781	caactctcgaatctctggtgttagagagcacagatgacagagctgacagacagccgtctcaaca	840
Db	2012	caactctcgaatctctggtgttagagagcacagatgacagagctgacagacagccgtctcaaca	2071
QY	841	ccgcacaacacagccgagagacacaagaagctccgcggaataatccgtgtttgagagctaaccc	900
Db	2072	ccgcacaacacagccgagagacacaagaagctccgcggaataatccgtgtttgagagctaaccc	2131
QY	901	ctgtgtccatctgcgatctccctctctctgcgaacatcacgaaaggggtgaagcgtgtgtg	960
Db	2132	ctgtgtccatctgcgatctccctctctctgcgaacatcacgaaaggggtgaagcgtgtgtg	2191
QY	961	aacgagggccatattgt	976
Db	2192	aacgagggccatattgt	2207
RESULT 2			
US-09-803-110-217			
: Sequence 217, Application US/09803110			
: GENERAL INFORMATION:			
: APPLICANT: Hinkle, Gregory J.			
: APPLICANT: Slater, Steven C.			
: TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof			
: FILE REFERENCE: 38-10115490.D			
: CURRENT APPLICATION NUMBER: US/09/803,110			
: PRIOR FILING DATE: 2001-03-12			
: PRIOR APPLICATION NUMBER: US 09/739,449			
: PRIOR FILING DATE: 2000-12-19			
: PRIOR APPLICATION NUMBER: US 09/514,000			
: PRIOR FILING DATE: 2000-02-23			
: PRIOR APPLICATION NUMBER: US 60/168,139			
: PRIOR FILING DATE: 1999-12-01			
: NUMBER OF SEQ ID NOS: 13351			
: SEQ ID NO 217			
: LENGTH: 397803			
: TYPE: DNA			
: ORGANISM: Agrobacterium tumefaciens			
: FEATURE:			
: NAME/KEY: unsure			

```

? LOCATION: (1)...(397803)
? OTHER INFORMATION: unsure at all n locations
US-09-803-110-217

Query Match          4.3%; Score 42; DB 6; Length 397803;
Best Local Similarity 60.5%; Pred. No. 2.5;
Matches 69; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

Oy 183 cgcgcgcggtgcgcgtatcatcaatagactcttgcgcgtctgtatccctcgatcac 242
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 66682 ccccgccatcatcgttcccaacttgcacaccccgcgcgcgagcagccctccacgacc 66741
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Oy 243 ccccttccacgcgcagccaagcgccgcgcctctctccgcgcgcgcgcgcgcgc 296
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 66742 ggaagtgtcgtcacccgcgcgcgtacttctctctccacgcgcgcgtctccaccgc 66795

RESULT 3
US-09-684-843A-1/c
: Sequence 1, Application US/09684843A
: GENERAL INFORMATION:
: APPLICANT: Rannum et al.
: TITLE OF INVENTION: SCA7 GENE AND METHODS OF USE
: FILE REFERENCE: Regents of the University of Minnesota
: CURRENT APPLICATION NUMBER: US/09/684,843A
: CURRENT FILING DATE: 2000-10-06
: PRIOR APPLICATION NUMBER: 60/056,170
: PRIOR FILING DATE: 1997-08-19
: PRIOR APPLICATION NUMBER: 09/135,934
: PRIOR FILING DATE: 1998-08-18
: NUMBER OF SEQ ID NOS: 14
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 477
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-684-843A-1

Query Match          4.1%; Score 40.4; DB 7; Length 477;
Best Local Similarity 53.9%; Pred. No. 1.7;
Matches 83; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

Oy 145 gcgcagcgcgaacccgggtgtgtgtctctccgcgcgaactgaacgcgcgagtagacgtatacc 204
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 334 gggcgggcggtggccgggtgctctccctccctggcggtggcgagcgcgctgctctct 275
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Oy 205 aatagactcttcgcgcgtcgtcatccctcgatcacccctgtttccagcgccagccaag 264
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 274 gctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 215
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Oy 265 gcagcgcgcctctctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 298
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 214 gctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 181
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 4
US-09-922-480-3
: Sequence 3, Application US/09922480
: GENERAL INFORMATION:
: APPLICANT: Sheppard, Paul O.
: TITLE OF INVENTION: SECRETED SALIVARY ZS163 POLYPEPTIDE
: FILE REFERENCE: 97-71
: CURRENT APPLICATION NUMBER: US/09/922,480
: CURRENT FILING DATE: 2001-08-03
: PRIOR APPLICATION NUMBER: US 60/124,820
: PRIOR FILING DATE: 1999-03-17
: NUMBER OF SEQ ID NOS: 9
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 3
: LENGTH: 657

```

```

; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Degenerate polynucleotide sequence for zsig63
; NAME/KEY: misc.feature
; LOCATION: (1)...(657)
; OTHER INFORMATION: n = A,T,C or G
US-09-922-480-3
```

```

Query Match 4.0%; Score 38.8; DB 7; Length 657;
Best Local Similarity 31.4%; Pred. No. 4.5;
Matches 89; Conservative 17; Mismatches 177; Indels 0; Gaps 0;
```

```

Oy 157 ccggggtgtgtctcctggcagctagcgagtagcagcatcacatcagcttct 216
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 340 ccnwmnmgttlytwsngcngcngcngcngcngcngcncnchnalhyngcngarcn 399
Oy 217 gccgtgtgtcatccctgtcatcacccctgtttcagcgccagcaagcgagcgct 276
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 400 gcnngcngcngcncnlytnacngcncncngcngcngcngcngcngcngcngcncn 459
Oy 277 cctgcgcgagcgatcggtctgagcgacgtgggagcaagatcattgtctaattgt 336
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 460 gtngcngcngarcncngcngcngarcngcngcngtngngcngcngcngcngcngcncn 519
Oy 337 aatcagctgggattaatgtaccccaaggagcgtctcagcactggagcgccgcaag 396
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 520 gtngcngcngarcncngcngcngarcngcngcngtngngcngcngcngcngcngarcn 579
Oy 397 acgaggtctacggcgttgagctccacagctaccccgccgagc 439
   :: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 580 wsnccngcngarcncngcngcngcngcngcngcngcngcngcngcngcngarc 622
```

```

RESULT 5
US-09-923-236-3
; Sequence 3, Application US/09923236
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: SECRETED SALIVARY ZSIG63 POLYPEPTIDE
; FILE REFERENCE: 97-71
; CURRENT APPLICATION NUMBER: US/09/923,236
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US 60/124,820
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 657
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Degenerate polynucleotide sequence for zsig63
; NAME/KEY: misc.feature
; LOCATION: (1)...(657)
; OTHER INFORMATION: n = A,T,C or G
US-09-923-236-3
```

```

Query Match 4.0%; Score 38.8; DB 7; Length 657;
Best Local Similarity 31.4%; Pred. No. 4.5;
Matches 89; Conservative 17; Mismatches 177; Indels 0; Gaps 0;
```

```

Oy 157 ccggggtgtgtctcctggcagctagcgagtagcagcatcacatcagcttct 216
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 340 ccnwmnmgttlytwsngcngcngcngcngcngcngcncnchnalhyngcngarcn 399
Oy 217 gccgtgtgtcatccctgtcatcacccctgtttcagcgccagcaagcgagcgct 276
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 400 gcnngcngcngcncnlytnacngcncncngcngcngcngcngcngcngcngcncn 459
```

```

Oy 277 cctgcgcgagcgatcggtcagcccgacgtgggagcaagatcattgtctaattgt 336
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 460 gtngcngcngarcncngcngcngarcncngcngtngngcngcngcngcngcngcncn 519
Oy 337 aatcagctgggattaatgtaccccaaggagcgtctcagcactggagcgccgcaag 396
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 520 gtngcngcngarcncngcngcngarcncngcngtngngcngcngcngcngcngcncn 579
Oy 397 acgaggtctacggcgttgagctccacagctaccccgccgagc 439
   :: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 580 wsnccngcngarcncngcngcngcngcngcngcngcngcngcngcngcngarc 622
```

```

RESULT 6
US-09-922-469-3
; Sequence 3, Application US/09922469
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: SECRETED SALIVARY ZSIG63 POLYPEPTIDE
; FILE REFERENCE: 97-71
; CURRENT APPLICATION NUMBER: US/09/922,469
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US 60/124,820
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 657
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Degenerate polynucleotide sequence for zsig63
; NAME/KEY: misc.feature
; LOCATION: (1)...(657)
; OTHER INFORMATION: n = A,T,C or G
US-09-922-469-3
```

```

Query Match 4.0%; Score 38.8; DB 7; Length 657;
Best Local Similarity 31.4%; Pred. No. 4.5;
Matches 89; Conservative 17; Mismatches 177; Indels 0; Gaps 0;
```

```

Oy 157 ccggggtgtgtctcctggcagctagcgagtagcagcatcacatcagcttct 216
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 340 ccnwmnmgttlytwsngcngcngcngcngcngcngcncnchnalhyngcngarcn 399
Oy 217 gccgtgtgtcatccctgtcatcacccctgtttcagcgccagcaagcgagcgct 276
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 400 gcnngcngcngcncnlytnacngcncncngcngcngcngcngcngcngcngcncn 459
Oy 277 cctgcgcgagcgatcggtcagcccgacgtgggagcaagatcattgtctaattgt 336
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 460 gtngcngcngarcncngcngcngarcncngcngtngngcngcngcngcngcngcncn 519
Oy 337 aatcagctgggattaatgtaccccaaggagcgtctcagcactggagcgccgcaag 396
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 520 gtngcngcngarcncngcngcngarcncngcngtngngcngcngcngcngcngarcn 579
Oy 397 acgaggtctacggcgttgagctccacagctaccccgccgagc 439
   :: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 580 wsnccngcngarcncngcngcngcngcngcngcngcngcngcngcngcngarc 622
```

```

RESULT 7
US-09-902-540-5862
; Sequence 5862, Application US/09902540
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
```



```

:   NUMBER OF SEQ ID NOS: 32581
:   SOFTWARE: FastSeq for Windows Version 4.0
:   SEQ ID NO 1232
:   LENGTH: 405
:   TYPE: DNA
:   ORGANISM: Lolium perenne
:   US-60-255-592-1232

Query Match          3.9%;   Score 37.8;   DB 8;   Length 405;
Best Local Similarity 46.2%;   Pred. No. 7.3;
Matches 126;   Conservative 0;   Mismatches 147;   Indels 0;   Gaps 0;

OY      153   gaaacggggtgtgtgtcctcggcagtcagcgcgcgagtcacagtcataccaataagac 212
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

[illegible]

```

GENERAL INFORMATION:
APPLICANT: Glenn, Matthew
APPLICANT: Norris, Michael G.
TITLE OF INVENTION: Compositions isolated from forage
FILE REFERENCE: 1056P
CURRENT APPLICATION NUMBER: US/60/253,654
CURRENT FILING DATE: 2000-11-28
NUMBER OF SEQ ID NOS: 32581
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 18403
LENGTH: 578
TYPE: DNA
ORGANISM: Festuca arundinaceae
US-60-253-654-18403

Query Match
Best Local Similarity      3.9%; Score 37.8; DB 8; Length 578;
Matches      57; Conservative    0; Mismatches   32; Indels     0; Gaps     0;
QY      209 cgactctgcgcgtgctgcatccctgatcaccocctgttttcagcgccagccaagcgag 268
          ||||| | |||| | | || | | || | | || | || || || || || || || || ||
Db      330 cgaagctgtctccgcgcgcctcgtcgtccaaacgccgacggcgtcgagcgccccgcgag 389
QY      269 ccgcgcctctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 297
          || || || || || || || || || || || || || || || || || || || || || ||
Db      390 ccacgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 418

RESULT      14
US-60-255-592-18403
: Sequence 18403, Application US/60255592
: GENERAL INFORMATION:
: APPLICANT: Glenn, Matthew
: APPLICANT: Norris, Michael G.
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 21, 2001, 00:34:56 : Search time 3479.37 Seconds
(without alignments)
80.020 Million cell updates/sec

Title: US-09-138-735-8
Perfect score: 18
Sequence: 1 tgcagcagcgcagagaagt 18

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl: *
1: gb_da1: *
2: gb_da2: *
3: gb_da3: *
4: gb_in1: *
5: gb_in2: *
6: gb_in3: *
7: gb_om: *
8: gb_ov: *
9: gb_pat1: *
10: gb_pat2: *
11: gb_ph: *
12: gb_pl1: *
13: gb_pl2: *
14: gb_pl3: *
15: gb_pl4: *
16: em_da1: *
17: em_da2: *
18: em_fun: *
19: em_htgo_hum: *
20: em_htgo_inv: *
21: em_htgo_rod: *
22: em_htg_hum1: *
23: em_htg_hum2: *
24: em_htg_hum3: *
25: em_htg_hum4: *
26: em_htg_hum5: *
27: em_htg_hum6: *
28: em_htg_hum7: *
29: em_htg_hum8: *
30: em_htg_inv1: *
31: em_htg_inv2: *
32: em_htg_other: *
33: em_htg_rod: *
34: em_hum1: *
35: em_hum2: *
36: em_hum3: *
37: em_hum4: *
38: em_hum5: *
39: em_hum6: *
40: em_hum7: *
41: em_in: *
42: em_om: *
43: em_or: *

44: em_ov: *
45: em_pat: *
46: em_ph: *
47: em_pl: *
48: em_rod: *
49: em_sts: *
50: em_sy: *
51: em_un: *
52: em_vl: *
53: gb_sts1: *
54: gb_sts2: *
55: gb_sts3: *
56: gb_sy: *
57: gb_un: *
58: gb_vl1: *
59: gb_vl2: *
60: gb_htg1: *
61: gb_htg2: *
62: gb_htg3: *
63: gb_htg4: *
64: gb_htg5: *
65: gb_htg6: *
66: gb_htg7: *
67: gb_htg8: *
68: gb_htg9: *
69: gb_htg10: *
70: gb_htg11: *
71: gb_htg12: *
72: gb_htg13: *
73: gb_htg14: *
74: gb_htg15: *
75: gb_htg16: *
76: gb_htg17: *
77: gb_htg18: *
78: gb_htg19: *
79: gb_htg20: *
80: gb_htg21: *
81: gb_htg22: *
82: gb_htg23: *
83: gb_htg24: *
84: gb_htg25: *
85: gb_pr1: *
86: gb_pr2: *
87: gb_pr3: *
88: gb_pr4: *
89: gb_pr5: *
90: gb_pr6: *
91: gb_pr7: *
92: gb_pr8: *
93: gb_pr9: *
94: gb_ro1: *
95: gb_ro2: *
96: gb_in4: *
97: gb_pr10: *
98: em_ba3: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	18 100.0	18 9	A48917	A48917 Sequence 8
2	18 100.0	18 9	AR047926	AR047926 Sequence
3	18 100.0	1589 10	AX086831	AX086831 Sequence
4	18 100.0	1589 93	HS0601826	AL136858 Homo sapi
5	18 100.0	3402 9	A48910	A48910 Sequence 1
6	18 100.0	3402 9	AR047920	AR047920 Sequence
7	18 100.0	3402 96	TC024190	U24190 Trypanosoma
8	18 100.0	11873 9	AR083116	AR083116 Sequence

9 18 100.0 11878 9 AR083115 Sequence
 10 18 100.0 11883 9 AR083114 Sequence
 11 17 94.4 915 77 AC085350
 12 17 94.4 920 71 AC034857
 13 17 94.4 1009 71 AC029282
 14 17 94.4 1671 53 AF309358
 15 17 94.4 4545 94 AB03044851
 16 17 94.4 5851 94 AF15253351
 17 17 94.4 6074 94 AF098866 Mus muscu
 18 17 94.4 65174 69 AC025214
 19 17 94.4 91470 12 AC011438
 20 17 94.4 110000 84 LMFCHR36_09
 21 17 94.4 137881 85 AC005184
 22 17 94.4 167590 80 AL356221
 23 17 94.4 171539 81 AL512788
 24 17 94.4 197871 73 AC068053
 25 17 94.4 200033 93 HSA251973
 26 16.4 91.1 160 54 HSPR3366
 27 16.4 91.1 652 2 BACPR1
 28 16.4 91.1 876 9 A14600
 29 16.4 91.1 876 10 E01244
 30 16.4 91.1 1399 94 AF192493
 31 16.4 91.1 1406 94 AF192494
 32 16.4 91.1 2573 94 AF190798
 33 16.4 91.1 3259 91 BC001222
 34 16.4 91.1 3436 89 AK001621
 35 16.4 91.1 5813 85 AB040944
 36 16.4 91.1 8014 63 AC014253
 37 16.4 91.1 14210 2 BS297025
 38 16.4 91.1 16427 65 AC020116
 39 16.4 91.1 36519 9 AR101859
 40 16.4 91.1 36929 78 AC091077
 41 16.4 91.1 48443 4 AC005656
 42 16.4 91.1 76452 65 AC020339
 43 16.4 91.1 78260 63 AC015439
 44 16.4 91.1 104467 61 AC010067
 45 16.4 91.1 106377 85 AC004132

ALIGNMENTS

RESULT 1
 LOCUS A48917 18 bp DNA PAT 07-MAR-1997
 DEFINITION Sequence 8 from Patent WO9605312.
 ACCESSION A48917
 VERSION A48917.1 GI:2302576
 KEYWORDS
 SOURCE unclassified.
 ORGANISM unclassified.
 REFERENCE 1 (bases 1 to 18)
 AUTHORS PARANHOS-BACCALA,G., LESENECHAL,M. and JOLIVET,M.
 TITLE NOVEL TRYPA NOSOMA CRUZI ANTIGEN, AND GENE CODING THEREFOR, THEIR
 JOURNAL APPLICATION TO THE DETECTION OF CHAGAS' DISEASE
 COMMENT Patent: WO 9605312-A 8 22-FEB-1996;
 BIO MERIEUX (FR)
 Other publication CA 2173957 960222
 Other publication AU 3169195 960307
 Other publication FR 2723589 960216.
 FEATURES
 source location/Qualifiers
 1..18
 /organism="unclassified"
 /db_xref="taxon:32644"
 BASE COUNT 5 a 4 c 7 g 2 t
 ORIGIN

Query Match 100.0%; Score 18; DB 9; Length 18;
 Best Local Similarity 100.0%; Pred. No. 5.8e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 tgcagcagcgagcaagt 18
 Db 1 TGCAGCAGCGCAGCAAGT 18

RESULT 2

LOCUS AR047926 18 bp DNA PAT 29-SEP-1999
 DEFINITION Sequence 8 from patent US 5820864.
 ACCESSION AR047926
 VERSION AR047926.1 GI:5970269
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 18)
 AUTHORS PARANHOS-BACCALA,G., LESENECHAL,M. and JOLIVET,M.
 TITLE TRYPA NOSOMA CRUZI ANTIGEN, GENE ENCODING THEREFOR AND METHODS OF
 DETECTING AND TREATING CHAGAS DISEASE
 JOURNAL Patent: US 5820864-A 8 13-OCT-1998;
 FEATURES location/Qualifiers
 1..18
 /organism="unknown"
 BASE COUNT 5 a 4 c 7 g 2 t
 ORIGIN

Query Match 100.0%; Score 18; DB 9; Length 18;
 Best Local Similarity 100.0%; Pred. No. 5.8e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 tgcagcagcgagcaagt 18
 Db 1 TGCAGCAGCGCAGCAAGT 18

RESULT 3

LOCUS AX086831 1589 bp DNA PAT 09-MAR-2001
 DEFINITION Sequence 783 from Patent WO0112659.
 ACCESSION AX086831
 VERSION AX086831.1 GI:13276044
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 REFERENCE 1 (bases 1 to 1589)
 AUTHORS Wiemann,S.
 TITLE Human dna sequences
 JOURNAL Patent: WO 0112659-A 783 22-FEB-2001;
 German Human Genome Project (DE)
 FEATURES location/Qualifiers
 1..1589
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 BASE COUNT 442 a 374 c 353 g 420 t
 ORIGIN

Query Match 100.0%; Score 18; DB 10; Length 1589;
 Best Local Similarity 100.0%; Pred. No. 2.6e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 tgcagcagcgagcaagt 18
 Db 316 TGCAGCAGCGCAGCAAGT 333

RESULT 4
 LOCUS HSM801826 1589 bp mRNA PRI 10-MAR-2001
 DEFINITION Homo sapiens mRNA, cDNA DKFZp434N2435 (from clone DKFZp434N2435);

complete cds.
 ACCESSION AL136858
 VERSION AL136858.1 GI:12053220
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 1589)
 AUTHORS Wiemann,S., Well,B., Wellenreuther,R., Gassenhuber,J., Glasl,S., Ansoerge,W., Boecher,M., Bloecker,H., Bauersachs,S., Blum,H., Labber,J., Duesterhoeft,A., Beyer,A., Koehner,K., Strack,N., Mewes,H.W., Olteneidei,B., Obermaier,B., Tampe,J., Heubner,D., Mambuti,R., Korn,B., Klein,M. and Poustka,A.
 Toward a Catalog of Human Genes and Proteins: Sequencing and Analysis of 500 Novel Complete Protein Coding Human cDNAs
 Genome Res. 11 (3), 422-435 (2001)
 JOURNAL 11230166
 PUBMED 2 (bases 1 to 1589)
 REFERENCE Poustka,A., Klein,M., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
 Direct Submission
 TITLE Submitted (18-JAN-2000) MIPS, Am Klopferspitz 18a, D-82152 Martinsried, GERMANY
 COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ), Email s.wiemann@dkfz-heidelberg.de; Heidelberg/Germany) within the cDNA sequencing consortium of the German Genome Project.
 This clone is available at the RZPD in Berlin.
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at http://www.mips.biochem.mpg.de/proj/cDNA/.
 location/Qualifiers
 1..1589
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="DKFZp434N2435"
 /tissue_type="testis"
 /clone_lib="434 (synonym: htest). Vector pSport1; host DH10B, sites NotI + SalI"
 /dev_stage="adult"
 78..1175
 /gene="DKFZp434N2435"
 78..1175
 /gene="DKFZp434N2435"
 /note="unknown protein"
 /product="hypothetical protein"
 /protein_id="CAB66792.1"
 /db_xref="GI:12053221"
 /translocation="MNV1P1A/PKGRICCECAEAERVCACATYTCGVYHOKA DMSIHKEICOLLIPRTSMPTFSEERHGOLOOOROKYILEFTYTTAOKTLPES KHEDAVPAALQSLERFKVLGLSSVEIYPAVPLAELSGRIVQAEVLFOKWTV LKSTDCSNATHSLHRLNGLIYAKKNVEARVHLANDIYFASCAFGTEIDRTSGYF HLANIFYDLKRLADTLTKVSEIHWAYENHNOVSOAHIOQMDLKGKFEEDTL DEAOEAERILRTSIILNIRESTSPKAPOKTFVLIYMLYLMNNSKQOYEMRL SLAKQQLDVHESTIOELSLISTEDHPIT"
 1561..1566
 polyA_signal
 polyA_site
 BASE COUNT 442 a 374 c 353 g 420 t
 ORIGIN
 Query Match 100.0%; Score 18; DB 93; Length 1589;
 Best Local Similarity 100.0%; Pred. No. 2.6e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 tgcacgacgagcagaagt 18
 ||||||||||||||||
 Db 316 TGCACGACGCGCAGAGT 333

RESULT 5
 A48910/c
 LOCUS A48910 3402 bp DNA
 DEFINITION Sequence 1 from Patent W09605312.
 ACCESSION A48910
 VERSION A48910.1 GI:2302570
 KEYWORDS
 SOURCE Trypanosoma cruzi.
 ORGANISM Trypanosoma cruzi
 Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae;
 Trypanosoma; Schizotrypanum.
 REFERENCE 1 (bases 1 to 3402)
 AUTHORS Paranhos-Baccala,G., Lesenechal,M. and Jolivet,M.
 NOVEL TRYPANOSOMA CRUZI ANTIGEN, AND GENE CODING THEREFOR; THEIR APPLICATION TO THE DETECTION OF CHAGAS' DISEASE
 Patent: WO 9605312-A 1 22-FEB-1996;
 BIO MERIEUX (FR)
 COMMENT Other publication CA 2173957 960222
 Other publication AU 3169195 960307
 Other publication FR 2723589 960216.
 location/Qualifiers
 1..3402
 /organism="Trypanosoma cruzi"
 /strain="G"
 /db_xref="taxon:5693"
 /dev_stage="EPIMASTIGOTE"
 BASE COUNT 889 a 818 c 958 g 737 t
 ORIGIN
 Query Match 100.0%; Score 18; DB 9; Length 3402;
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0;
 QY 1 tgcacgacgagcagaagt 18
 ||||||||||||||||
 Db 1459 TGCACGACGCGCAGAGT 1442
 RESULT 6
 AR047920/c
 LOCUS AR047920 3402 bp DNA
 DEFINITION Sequence 1 from patent US 5820864.
 ACCESSION AR047920
 VERSION AR047920.1 GI:5970263
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 3402)
 AUTHORS Paranhos-Baccala,G., Lesenechal,M. and Jolivet,M.
 TITLE Trypanosoma cruzi antigen, gene encoding therefor and methods of detecting and treating chagas disease
 Patent: US 5820864-A 1 13-OCT-1998;
 JOURNAL Location/Qualifiers
 1..3402
 /organism="unknown"
 BASE COUNT 888 a 821 c 956 g 737 t
 ORIGIN
 Query Match 100.0%; Score 18; DB 9; Length 3402;
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0;
 QY 1 tgcacgacgagcagaagt 18
 ||||||||||||||||
 Db 1459 TGCACGACGCGCAGAGT 1442
 RESULT 7
 TC024190/c
 LOCUS TC024190 3402 bp mRNA
 INW 04-AUG-1997

DEFINITION Trypanosoma cruzi Tc40 antigen (Tc40) mRNA, complete cds.
ACCESSION U24190
VERSION U24190.1 GI:790645
KEYWORDS
SOURCE Trypanosoma cruzi.
ORGANISM Trypanosoma cruzi.
REFERENCE 1 (bases 1 to 3402)
AUTHORS Lesnechal, M., Duret, L., Cano, M.I., Mortara, R.A., Jolivet, M., Camargo, M.E., da Silveira, J.F. and Paranhos-Baccala, G.
TITLE Cloning and characterization of a gene encoding a novel immunodominant antigen of Trypanosoma cruzi
JOURNAL Mol. Biochem. Parasitol. 87 (2), 193-204 (1997)
MEDLINE 97391123
REFERENCE 2 (bases 1 to 3402)
AUTHORS Lesnechal, M., Franco Da Silveira, J., Mortara, R.A., Duret, L., Camargo, M.E., Jolivet, M. and Paranhos-Baccala, G.
TITLE Direct Submission
JOURNAL Submitted (06-APR-1995) Mylene Lesnechal, Biomerieux, 22 rue Saint Jean de Dieu, Lyon 69007, France
FEATURES
source 1..3402
/organism="Trypanosoma cruzi"
/strain="G"
/db_xref="taxon:5693"
/dev_stage="epimastigote"
1..35
/gene="Tc40"
/note="spliced leader"
1..3402
/gene="Tc40"
266..3013
/gene="Tc40"
/codon_start=1
/product="Tc40 antigen"
/protein_id="AAC47657.1"
/db_xref="GI:790646"
/translation="MTVTVDLFNNAKPSNNEGRWVSVDATENEVEAQRVLADSOFY
LAYTKRRHVLRYKRSNLLKGTVAHSPHIAVKVNRNVNVAASAGGEFFVAVT
DETEASNGKPDLAARLVYFELQDPVIPCSPFINKESORPDLLVLYEQOALID
SSLIREFPVESELEAIONCTLRITQDPVSNSTCSVSGSGMFEFTPEPMVACT
LNRSRSPACCEGEPYKALHLDATVEENSVLAASKGYOHLITVAEPNLARK
EYIDGSIYVMESSRETFVFDKQALVNMHSPHFTCTHTMPCQVORNGCFENRT
AGSCVLAQMSNRLITFLHRSRREDOGOKTSVATKPGVSGTDAASSHTNT
TAAAPSPAPSPVSAAPAKAAPPAARSAEPHSGKTIANLVOLGINTORSVSTG
APATTSTAVTSTTAPORTSPYGHNGRPVTAAGLVANVGASAASSPTAAKPTGEX
ASACETSSVAIINATPRPALHNASIPQAPDGVLAAYVSGEVSRLSRISVITNT
SVYKILPTPTIRDEQLNLGLLEAOMTELQOSRPPTOPRDTSSAKSSVEFTYLV
LIADLSRNTITKGVKRGVNEAIVMLHLDHEVRHATIGRLKQTKNIIKSLDELKEST
TQFTAQLTQTVENLVKRELAEVIGSINGSLIVKRNASLOKELNIMSGVLDKMR
MREELCTLRSAKRAKATMPDSSLIHATSSFOGRSAPETILATLSMVEEOYROELE
VYLMQOPELILFLFSLITRENENAVSELIENVEETNDVYCSVLLOLILAAATEAKE
VYVGVAIDILSERDQIAONGALGSKITMTAMRAFERARSETTSRSLQCLKNLEKILQ
S"

BASE COUNT 889 a 818 c 958 g 737 t
ORIGIN

Query Match 100.0%; Score 18; DB 96; Length 3402;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgcagcagcgagcagaagt 18
|||||
Db 1459 TGCAGCAGCGCAGAGT 1442

RESULT 8
LOCUS AR083116 11873 bp DNA
DEFINITION Sequence 32 from patent US 5976803.
ACCESSION AR083116 PAT 01-SEP-2000

VERSION AR083116.1 GI:10009906
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 11873)
AUTHORS Week, K.D.
TITLE Genetic test for equine severe combined immunodeficiency disease
JOURNAL Patent: US 5976803-A 32 02-NOV-1999;
FEATURES
source 1..11873
/organism="unknown"
BASE COUNT 3585 a 2320 c 2669 g 3299 t
ORIGIN

Query Match 100.0%; Score 18; DB 9; Length 11873;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgcagcagcgagcagaagt 18
|||||
Db 6429 TGCAGCAGCGCAGAGT 6446

RESULT 9
LOCUS AR083115 11878 bp DNA
DEFINITION Sequence 31 from patent US 5976803.
ACCESSION AR083115 PAT 01-SEP-2000
VERSION AR083115.1 GI:10009905
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 11878)
AUTHORS Week, K.D.
TITLE Genetic test for equine severe combined immunodeficiency disease
JOURNAL Patent: US 5976803-A 31 02-NOV-1999;
FEATURES
source 1..11878
/organism="unknown"
BASE COUNT 3586 a 2323 c 2668 g 3301 t
ORIGIN

Query Match 100.0%; Score 18; DB 9; Length 11878;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgcagcagcgagcagaagt 18
|||||
Db 6429 TGCAGCAGCGCAGAGT 6446

RESULT 10
LOCUS AR083114 11883 bp DNA
DEFINITION Sequence 28 from patent US 5976803.
ACCESSION AR083114 PAT 01-SEP-2000
VERSION AR083114.1 GI:10009904
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 11883)
AUTHORS Week, K.D.
TITLE Genetic test for equine severe combined immunodeficiency disease
JOURNAL Patent: US 5976803-A 28 02-NOV-1999;
FEATURES
source 1..11883
/organism="unknown"
BASE COUNT 3588 a 2324 c 2669 g 3302 t

ORIGIN

Query Match 100.0%; Score 18; DB 9; Length 11883;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgcagcagcgagcagaag 18
|||||

Db 6432 TGCACGACGCGCAGAGT 6449

RESULT 11
AC085350
LOCUS
DEFINITION Giardia intestinalis clone KJ5074 strain WB-C6, LOW-PASS SEQUENCE
SAMPLING.
AC085350.1 GI:11465342
VERSION
KEYWORDS HTG; HTGS_PHASE0.
SOURCE Giardia intestinalis.
ORGANISM Giardia intestinalis.
REFERENCE Eukaryota; Diplomonadida; Hexamitidae; Giardinae; Giardia.
AUTHORS McArthur,A.G., Morrison,H.G., Nixon,J.E.J., Passamaneck,N.O.,E., Kim,U., Hinkle,G., Crocker,M.K., Holder,M.E., Farr,R., Reich,C.I., Olsen,G.J., Aley,S.B., Adam,R.D., Gillin,F.D. and Sogin,M.L.
TITLE The Giardia genome project database
JOURNAL PEMS Microbiol. Lett. 189 (2), 271-273 (2000)
MEDLINE 20389616
REFERENCE 2 (bases 1 to 915)
AUTHORS Kim,U., Morrison,H.G., McArthur,A.G., Nixon,J., Eakin,N.O., Fierro,L.A., Aley,S.B. and Sogin,M.L.
TITLE Direct Submission
JOURNAL Submitted (30-NOV-2000) Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Marine Biological Laboratory, 7 MBL Street, Woods Hole, MA 02543-1015, USA
COMMENT * NOTE: This record contains 1 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

FEATURES
source
1. 915
/organism="Giardia intestinalis"
/strain="WB-C6"
/db_xref="taxon:5741"
/clone="KJ5074"
BASE COUNT 278 a 195 c 274 g 166 t 2 others
ORIGIN

Query Match 94.4%; Score 17; DB 77; Length 915;
Best Local Similarity 100.0%; Pred. No. 8.5e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgcagcagcgagcagaag 17
|||||

Db 818 TGCACGACGCGCAGAG 834

RESULT 12
AC034857/C
LOCUS
DEFINITION Giardia intestinalis clone AJ1539 strain WB-C6, LOW-PASS SEQUENCE
SAMPLING.

AC034857
VERSION
KEYWORDS HTG; HTGS_PHASE0.
SOURCE Giardia intestinalis.
ORGANISM Giardia intestinalis.
REFERENCE Eukaryota; Diplomonadida; Hexamitidae; Giardinae; Giardia.
AUTHORS Morrison,H.G., McArthur,A.G., Nixon,J., Eakin,N.O., Kim,U., Crocker,M.C., Hinkle,G., Holder,M.E. and Sogin,M.L.
TITLE Giardia: a model for ancient eukaryotic genome analysis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 920)
AUTHORS McArthur,A.G., Morrison,H.G., Nixon,J., Eakin,N.O., Kim,U., Crocker,M.C., Hinkle,G., Holder,M.E. and Sogin,M.L.
TITLE Direct Submission
JOURNAL Submitted (05-APR-2000) Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Marine Biological Laboratory, 7 MBL Street, Woods Hole, MA 02543-1015, USA
COMMENT * NOTE: This record contains 1 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

FEATURES
source
1. 920
/organism="Giardia intestinalis"
/strain="WB-C6"
/db_xref="taxon:5741"
/clone="AJ1539"
BASE COUNT 187 a 250 c 223 g 260 t
ORIGIN

Query Match 94.4%; Score 17; DB 71; Length 920;
Best Local Similarity 100.0%; Pred. No. 8.5e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgcagcagcgagcagaag 17
|||||

Db 678 TGCACGACGCGCAGAG 662

RESULT 13
AC029282
LOCUS
DEFINITION Giardia intestinalis clone HF2663 strain WB-C6, LOW-PASS SEQUENCE
SAMPLING.
AC029282.1 GI:7384133
VERSION
KEYWORDS HTG; HTGS_PHASE0.
SOURCE Giardia intestinalis.
ORGANISM Giardia intestinalis.
REFERENCE Eukaryota; Diplomonadida; Hexamitidae; Giardinae; Giardia.
AUTHORS Morrison,H.G., McArthur,A.G., Nixon,J., Eakin,N.O., Kim,U., Crocker,M.C., Hinkle,G., Holder,M.E. and Sogin,M.L.
TITLE Giardia: a model for ancient eukaryotic genome analysis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1009)
AUTHORS Kim,U., Crocker,M.K., Holder,M.E. and Sogin,M.L.
TITLE Direct Submission
JOURNAL Submitted (01-APR-2000) Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Marine Biological Laboratory, 7 MBL Street, Woods Hole, MA 02543-1015, USA
COMMENT * NOTE: This record contains 1 individual

* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

FEATURES
source 1009: contig of 1009 bp in length.
1.1009
Location/Qualifiers

BASE COUNT 299 a 211 c 290 g 207 t 2 others
ORIGIN
/organism="Giardia intestinalis"
/strain="WB-C6"
/db_xref="taxon:5741"
/clone="HF2663"

Query Match 94.4%; Score 17; DB 71; Length 1009;
Best Local Similarity 100.0%; Pred. No. 8.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 tgcagcagcgccagaag 17
|||||
Db 766 TGCAGCAGCGCAGAG 782

RESULT 14
AF309358 1671 bp DNA STS 30-MAR-2001
LOCUS Hordeum vulgare STS marker ABG452, sequence tagged site.
DEFINITION AF309358
ACCESSION AF309358.1 GI:13383483
KEYWORDS STS.
SOURCE barley.
ORGANISM Hordeum vulgare
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae; Triticeae;
Hordeum.

REFERENCE 1 (bases 1 to 1671)
AUTHORS Collins,N.C., Lahaye,T., Peterhansel,C., Freialdenhoven,A.,
Corbitt,M. and Schulze-Lefert,P.
TITLE Sequence haplotypes revealed by sequence-tagged site fine mapping
of the ror1 gene in the centromeric region of barley chromosome 1h
JOURNAL Plant Physiol. 125 (3), 1236-1247 (2001)
MEDLINE 21140309
PUBMED 11244105

REFERENCE 2 (bases 1 to 1671)
AUTHORS Collins,N.C., Lahaye,T., Peterhansel,C., Freialdenhoven,A.,
Corbitt,M. and Schulze-Lefert,P.
TITLE Direct Submission

JOURNAL Submitted (28-SEP-2000) Sainsbury Laboratory, John Innes Institute,
Colney Lane, Norwich, Norfolk NR4 7UH, UK
COMMENT primers: 5' segment: forward GCGAGTGTGATATGTGGCATC, reverse
ATATGACAGACGCTCTCATCC; 3' segment: forward
TCTACCTGCTCTTCAAA, reverse AGCTCCACCAACCCACAGTG.
FEATURES
source 1.1671
Location/Qualifiers

1.1671
/organism="Hordeum vulgare"
/db_xref="taxon:4513"
/chromosome="1H"
/map="1H centromeric region"
/note="genotype: BCIngrid mlo-5"
STS
primer_bind 1098..1117
STS primer_bind 1098..11671
BASE COUNT 464 a 408 c 427 g 372 t
ORIGIN

Query Match 94.4%; Score 17; DB 53; Length 1671;
Best Local Similarity 100.0%; Pred. No. 7.7e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 tgcagcagcgccagaag 17
|||||
Db 265 TGCAGCAGCGCAGAG 281

RESULT 15
AB030448S1/c

LOCUS AB030448S1 4545 bp DNA 14-APR-2000
DEFINITION Mus musculus V1bR gene for V1b arginine vasopressin receptor, exon
1.

ACCESSION AB030448
VERSION AB030448.1 GI:6683122
KEYWORDS V1bR; V1b arginine vasopressin receptor.
SEGMENT 1 of 2
SOURCE Mus musculus (strain:129SVJ) 4-8 weeks female tissue lib: liver DNA.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (sites)
AUTHORS Kikuchi,S., Tanoue,A., Goda,N., Matsuo,N. and Tsujimoto,G.
TITLE Structure and sequence of the mouse V1a and V1b vasopressin
receptor genes
Jpn.J. Pharmacol. 81 (4), 388-392 (1999)
MEDLINE 20132529
REFERENCE 2 (bases 1 to 4545)
AUTHORS Tsujimoto,G. and Kikuchi,S.
TITLE Direct Submission

JOURNAL Submitted (27-JUL-1999) to the DDBJ/EMBL/Genbank databases. Gozoh
Tsujimoto, National Children's Medical Research Center, Molecular
Pharmacology, 3-35-31 Taitshido, Setagaya-ku, Tokyo-to 154-8509,
Japan (E-mail:skikuchi@nmc.go.jp, Tel:81-3-3414-8121(ex.2733),
Fax:81-3-3419-1252)
FEATURES
source 1.4545
Location/Qualifiers

1.4545
/organism="Mus musculus"
/strain="129SVJ"
/db_xref="taxon:10090"
/dev_stage="4-8 weeks"
/sex="female"
/tissue lib="liver"
1424..2354
/gene="V1bR"
/note="CDS is reported in Acc# AB030449"
/number=1
/product="V1b arginine vasopressin receptor"

BASE COUNT 1012 a 1246 c 1013 g 1274 t
ORIGIN

Query Match 94.4%; Score 17; DB 94; Length 4545;
Best Local Similarity 100.0%; Pred. No. 6.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 tgcagcagcgccagaag 17
|||||
Db 2185 TGCAGCAGCGCAGAG 2169

Search completed: September 21, 2001, 00:35:06
Job time: 29561 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2001, 16:22:25 ; Search time 3479.37 Seconds
(without alignments)
2640.663 Million cell updates/sec

Title: US-09-138-735-1_COPY_1232_1825
Perfect score: 594
Sequence: 1 caggtacagcgtacgctt.....aaacgagctccgtgcgata 594

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_on:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_p11:*
13: gb_p12:*
14: gb_p13:*
15: gb_p14:*
16: gb_ba1:*
17: gb_ba2:*
18: gb_fun:*
19: gb_hum:*
20: gb_hum2:*
21: gb_hum3:*
22: gb_hum4:*
23: gb_hum5:*
24: gb_hum6:*
25: gb_hum7:*
26: gb_hum8:*
27: gb_hum9:*
28: gb_hum10:*
29: gb_hum11:*
30: gb_hum12:*
31: gb_hum13:*
32: gb_hum14:*
33: gb_hum15:*
34: gb_hum16:*
35: gb_hum17:*
36: gb_hum18:*
37: gb_hum19:*
38: gb_hum20:*
39: gb_hum21:*
40: gb_hum22:*
41: gb_hum23:*
42: gb_hum24:*
43: gb_hum25:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	594	100.0	3402	9	AR047920
2	590.8	99.5	3402	9	A48910
3	590.8	99.5	3402	96	TCU24190
4	45	7.6	47852	3	MTW023
5	44	7.4	2402	3	MIORHO
6	43.2	7.3	2028	15	SSI132828
7	43.2	7.3	158390	60	AC008532
8	42.8	7.2	767	6	ETIMAX

C	9	42.2	7.2	14.0812	13	AP001081	AP001081	Oryza sativa
C	10	41.6	7.0	5880	55	HSEA	D14486	Equine herpesvirus 1
C	11	41.2	6.9	2479	95	RNUB79660	U87960	Rattus norvegicus
C	12	40.4	6.8	477	10	E34519	E34519	SCA1 gene
C	13	40.4	6.8	477	88	AF020275	AF020275	Homo sapiens
C	14	40.4	6.8	3932	93	HSMB01297	AL133030	Homo sapiens
C	15	40.4	6.8	3941	93	RNLAR2	X83546	R. norvegicus
C	16	40.4	6.8	5477	85	AB051453	AB051453	Homo sapiens
C	17	40.4	6.8	65350	3	MTV004	AL009198	Mycobacterium tuberculosis
C	18	40.4	6.8	143618	91	AP000056	AP000056	Homo sapiens
C	19	40.4	6.8	150036	91	AP000057	AP000057	Homo sapiens
C	20	40.4	6.8	157086	91	AP000052	AP000052	Homo sapiens
C	21	40.4	6.8	166447	87	AC018751	AC018751	Homo sapiens
C	22	40.4	6.8	163237	87	AC009516	AC009516	Homo sapiens
C	23	40.4	6.8	171350	72	AC058816	AC058816	Homo sapiens
C	24	40.4	6.8	176051	88	AC023490	AC023490	Homo sapiens
C	25	40.4	6.8	213656	86	AC007957	AC007957	Homo sapiens
C	26	40.2	6.8	701	13	AF310215	AF310215	Sorghum bicolor
C	27	40.2	6.8	138273	83	AP003297	AP003297	Oryza sativa
C	28	40.2	6.8	172307	71	AC044842	AC044842	Homo sapiens
C	29	40	6.7	624	5	AF139019	AF139019	Cepaea nemoralis
C	30	40	6.7	122395	78	AC090433	AC090433	Chlamydomonas reinhardtii
C	31	40	6.7	130569	78	AC090650	X17207	Arabidopsis thaliana
C	32	39.6	6.7	618	14	CRCGCRI	X17207	Chlamydomonas reinhardtii
C	33	39.6	6.6	947	53	CNS0794Y	AL434840	T3 end of Arabidopsis thaliana
C	34	39.4	6.6	1442	88	AF087653	AF087653	Homo sapiens
C	35	39.2	6.6	818	8	CHKPOTAMI	L87113	Gallus gallus
C	36	39.2	6.6	1528	8	CHKPOTAMN	M8100	Gallus gallus
C	37	39.2	6.6	2771	94	MMZINCPR	X9550	Gallus gallus
C	38	39.2	6.6	135382	94	MMZINCPR	X95504	M. musculus
C	39	39.2	6.6	156589	92	HS537K23	AL034405	Human DNA
C	40	39	6.6	259	85	AB018491	AB018491	Homo sapiens
C	41	38.8	6.5	720	85	AB007820	AB007820	Homo sapiens
C	42	38.8	6.5	1030	14	HVCAB2	X12735	Barley
C	43	38.6	6.5	2858	97	MEAPO4A4	X68361	M. fascicularia
C	44	38.6	6.5	54327	78	AC090433	AC090433	Chlamydomonas reinhardtii
C	45	38.6	6.5	86030	78	AC090436	AC090436	Chlamydomonas reinhardtii

ALIGNMENTS

RESULT	1						
AR047920		AR047920	3402 bp	DNA	PAT	29-SEP-1999	
LOCUS		Sequence	1 from patent US	5820864.			
DEFINITION		AR047920					
ACCESSION		AR047920.1	GI:5970263				
VERSION							
KEYWORDS		.					
SOURCE		Unknown.					
ORGANISM		Unknown.					
REFERENCE		Unclassified.					
AUTHORS		1 (bases 1 to 3402)					
TITLE		Parainosoma baccali, G., Lesenechal, M. and Jollivet, N. Trypanosoma cruzi antigen, gene encoding therefor and methods of detecting and treating Chagas disease					
JOURNAL		Patent; US 5820864-A 1 13-OCT-1998;					
FEATURES		Location/Qualifiers					
SOURCE		1..3402					
BASE COUNT		/organism="unknown"					
ORIGIN		888 a 821 c 956 g 737 t					

Query Match	Score	DB	Length
100.0%	594	9	3402

Matches 594; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 caggtacagcgtaacggtcttctgctcaatcggtacagccgacggtagctgctcctgct 60

1232 CAGGTACAGCGTACGGCTTTTGGCTCAATCGTACAGCCGACGGTAGCTGCGTCCTGGCT 1291

61 gacatgtcgaattcgattgacgatacttccatctccggtcctccgcagggagaacagcag 12

[illegible]

	RESULT	2			
	A48910				
	LOCUS	A48910	3402 bp	DNA	
	DEFINITION	Sequence 1 from patent WO9605312.		PAT	07-MAR-1997
	ACCESSION	A48910			
	VERSION	A48910.1	GI:2302570		
	KEYWORDS				
	SOURCE				
	ORGANISM	Trypanosoma cruzi. Eukaryota; Euzoenzoa; Kinetoplastida; Trypanosomatidae; Trypanosoma; Schizotrypanum. 1 (bases 1 to 3402) Paranhos-Baccala,G., Lesenechal,M. and Jollivet,M. NOVEL TRYPANOSOMA CRUZI ANTIGEN, AND GENE CODING THEREFOR; THEIR APPLICATION TO THE DETECTION OF CHAGAS' DISEASE Patent: WO 9605312-A 1 22-FEB-1996; BIO MERIEUX (FR) Other publication CA 2173957 960222 Other publication AU 3169195 960307 Other publication FR 2723589 960216.			
	COMMENT				
	JOURNAL				
	FEATURES				
	Source	1..3402			
		/organism="Trypanosoma cruzi" /strain="G" /db_xref="taxon:5693" /dev_stage="EPIMASTIGOTE"			
		Location/Qualifiers			
		889 a 818 c 958 g 737 t			
	BASE COUNT				
	ORIGIN				

Query Match	99.58;	Score 590.8;	DB 9;	Length 3402;
-------------	--------	--------------	-------	--------------

Matches	592;	Conservative	0;	Mismatches	2;	Indels	0;	Gaps	0;
---------	------	--------------	----	------------	----	--------	----	------	----

1 caggtacagcgtaacggtttgtcttcaatcgtacagccgacggtagctgtgctcctgct 60

b 1232 CAGGTACAGCGTAACGGCTTTGCTTCAATCGTACAGCCGACGGTAGCTGCGTCTGGCT 1291

[illegible]

```

CDS
266..3013
/gene="Tc40"
/codon_start=1
/product="Tc40 antigen"
/protein_id="AAC47657.1"
/db_xref="GI:790646"
/translation="MTVTVDFNFNAKPSNNNGRWSDAATFNEPVEAIVLADSOFY
LAYMKRRHYLRVKKRSNLLKGTVRASKPLPAKFPVNSVNAASGKGEPFWVYT
DEPASKGPDLLARLVKKYFYFKLDQDPVTLPFPAIESORPDLVLLETQAILID
SSSLIERFDVESLEATLQRMCTLTLLTLPQVSENSLICSVSGSEFTTLEPTMAACT
LRNSSTVSMACCCEGEPKVALKLDATVEENSVYLAASTHYKQWMLTGAEPMLIKR
FVIGDSIVAMSSRETPAIVEDRKKQLALVNMHSGVNSCTHPMPCOVORGFENRTF
ADGCSVLADMSNRLTIFHLSSRRREPOGQSVVATRKQGVCSGTDAASSHTMTN
TSAASAASPSPVSAAPAKAAAPPAASAABEPHSKTIIANLVNGLINVTORSVSTG
APATRSSTAVYSTTAPQRTSPYGHNRPTVAGLVANSGSAASSTPLAAKPTGEK
ASACETSSVAINATRALNHAASLPQPTDVLAAAYQSGEWHSLERLESYITNTN
SRVRLKLPDTRIRDEHQLNLNGLEAOWTEILOOSPPTQOTPRONKSIKSRIDEALKEST
LIADLSRNITKGVKRGVNEAIMLHLDEYHVALIGNRLROKNI IKSRIDEALKEST
TQPTFAOLITQYEMENVKREIAEVGSLINGSITLVYKNAAISLOKELNISTMSGVLDERR
MRBELCTLRESVARKKATMPDSSLIHATISSRQGRSAPEITLATLSMVRQOYRGGLE
YMLDAQPSLLLRFLSLITRENNAYSELLIENVTPEMDWCSTVLQDILEAAVEAREE
VVMVAQIDILISERDQIAQNGALSGSKLTITNMAFERAORSETTSRSFLOCLKNLEKLO
S"
BASE COUNT      889 a      818 c      958 g      737 t
ORIGIN
Query Match      99.5%; Score 590.8; DB 96; Length 3402;
Best Local Similarity 99.7%; Pred. No. 6.9e-114;
Matches 592; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY      1      cagtgacagcgtaacgcgtttgtgtcttaacgtacagccgacgtagtgcgtctgcgt 60
Db      1232  CAGGTACAGCGGTAAACGGGTTTGTGTTCAATGATACAGCGGAGGTAGCTGCTGGACT 1291
OY      61      gacatgctgattgaattagacatcttcacatctccggtctctccggaagggaagaacag 120
Db      1232  GACATGTGGAATCGATTACGATTCATCTCCAGTCTCCCGGAGGGAAGAACAGAG 1351
OY      121     cgaagccaaanaaacctcgttagtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 180
Db      1352  CGAGCCCAAAAACATCGTGGTAGTGGCGAGACCGGGAACCGGGGTGTGTCTCCGGGCACT 1411
OY      181     gaecgcgcgagtagcagtcataccaatacagactctgcgcgctgtgtcaatccctgcataca 240
Db      1412  GACGCGGCGAGTACGATCATACCAATACGACTTCTCGCGGCTGTCATCCCTGCATCA 1471
OY      241     cccctgtttaaagcccaagcgcaacgcgcgcgtctctgcgcgcgcgcgcgcgcgcgcgcgc 300
Db      1472  CCCCCTGTTTACGCCCGCAGGAAGGCACCGCGCTCTCGCGCGCGCGCGAGTGGCTGAG 1531
OY      301     ccgcacgtggggaagaatcatctgaatctagtgaatcaagcttgaggatgaatgaac 360
Db      1532  CCGCACGTGGGGAGACCAATCTTCTCTAATCTAGTGAATCAAGCTGGGATTAATGTCAAC 1591
OY      361     caaagagcgctgcagcactgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 420
Db      1592  CAAAGGACGCTGTCAGACACTGAGACGCGCGCGCACGAGAGGTCTACGGCGGTGACGTC 1651
OY      421     acgactaccgcccgagcgcgaacaaagttccatacggcgcaaatgycgcgactgtgaagct 480
Db      1652  ACGACTACCGCCCCCGACGAAACAAGTCCATACGGGGCAATATGGCGACCTGTGACGGCT 1711
OY      481     ggattggcgagcactaatagtggtgcagcgcgcgctcgtctcccaagcccgcgcgga 540
Db      1712  GGATTGGTGGCGACCTAATATAGTGTGTCACGCGCGGCTGTCTCCACAGCGCGGCGAA 1771
OY      541     ccaacaggaagaagaagcctccgcgcgcgcatgtgaacagagctccgttgcgata 594
Db      1772  CCAACAGGAGAAGAAGAGCGCTCCGCGGCATGTGAAGAGAGCTCGGTGCGATA 1825

```

MTV023 47852 bp DNA BCT 17-JUN-1998
LOCUS MTV023 47852 bp DNA BCT 17-JUN-1998
DEFINITION Mycobacterium tuberculosis H37RV complete genome; segment 148/162.
ACCESSION AL022022 AL123456
VERSION AL022022.1 GI:3261554
KEYWORDS
SOURCE Mycobacterium tuberculosis.
ORGANISM Mycobacterium tuberculosis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium tuberculosis complex.
1 (bases 1 to 47852)
Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C.,
Harris, D., Gordon, S.V., Eigmeier, K., Gas, S., Barry III, C.E.,
Tekaita, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T.,
Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S.,
Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J.,
Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,
Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S.,
Squares, S., Squares, R., Sulston, J.E., Taylor, K., Whitehead, S. and
Barrell, B.G.
Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence
Nature 393 (6685), 537-544 (1998)
98295987
Erratum: [[published erratum appears in Nature 1998 Nov
12:396(6707):190]]
2 (bases 1 to 47852)
Parkhill, J.
Direct Submission
Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
Trust Genome Campus, Hinxton, Cambridge CB10 1SA unite de Genetique
Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
On Jun 27, 1998 this sequence version replaced gi:2924430.
Notes:
Details of M. tuberculosis sequencing at the Sanger Centre are
available on the World Wide Web.
(URL: http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have
been renumbered from the original cosmid submissions but the old
gene designations are in brackets after the new gene numbers.
Gene prediction was based on a Hidden Markov Model of TB genes
implemented in Tbpase (Krogh) supplemented with visual inspection
of positional base preference in codons, especially where there is
an increase in the observed/expected third position G + C.
CAUTION: In some cases we may not have predicted the correct
initiation codon. Where possible we choose an initiation codon
(atg, ggg, or ttg) which is preceded by an upstream ribosome
binding site sequence (optimally 5-13bp before the initiation
codon). If this cannot be identified we choose the most upstream
initiation codon.
Location/Qualifiers
1. 47852
/organism="Mycobacterium tuberculosis"
/strain="H37RV"
/db_xref="taxon:1773"
1. 18
/organism="Mycobacterium tuberculosis"
/strain="H37RV"
/db_xref="taxon:1773"
/clone="Y13E12"
1. 47852
/note="fragment designated v023. Does not represent a
physical clone"
complement(1..1695)
/gene="RV3494c"
complement(1..1695)
/gene="RV3494c"
complement(1..1695)
/gene="RV3494c"
complement(1..1695)
/note="RV3494c" (MTV023.01c), len: 564. Unknown Pro-rich
protein similar to several Mycobacterium tuberculosis
proteins e.g. MTC128.14 (515 aa). MTCY19H5.28c (516 aa) and
(MTV051.09). Has hydrophobic stretch, possibly signal

peptide at N-terminus. FASTA scores: 297050|MTC128_14
(515 aa) opt: 1006 z-score: 582.8 E(): 5.9e-25; 34.38
identity in 559 aa overlap: 297182|MTCY19H5_28 (516 aa)
opt: 979 z-score: 567.7 E(): 4.1e-24; 33.58 identity in
555 aa overlap. Tbpase score is 0.897"
/codon_start=1
/transl_table=11
/product="hypothetical protein RV3494c"
/protein_id="CAI17731.1"
/db_xref="GI:2924431"
/db_xref="SPTREMBL:O53539"
/translation="MIDRLAKIOLSTIFANITVTITLSVMAIFIRLPATRGICITGVSA
DEVAGGLTKMNAVITYRGVAVGVESVGLPNNGVIAHMLNSTALPSNVTAIVKSVS
ALIEQYIDLVPPENPSTKLRNGFRIQONTIRIGDVAOLLRQATLLSGLDPTRLRE
LHEAFATNAGAPELARLIESARLIVDEANANYPQVSLIDAGFPLOAQIRAGGDI
KSLADGARTNOLRAADPRRLTADAPDAIDAEATARGISRPSPALAAISLANGR
VGVYHKSIEDOLLVLPAPATITTSAGVPODEGAKDKFDLIDHPPCMGFLPPLP
LVNSPDESREIRPMYCKTQNDSTYRGANRYCOEPFGKRAATVYLCRDPRRY
PVGTNWRGPPPIYGVETDGRNIIPLPNKFPYIIPCADDPGVPIVGPVGVAGVG
PAPHQPAQAPAPPPNDGPPPPPTSMPPGVPVPPVYPATIIPPPPPGPGPPGP
APGPPOASGPAITVDOLSGAFADPAGGTGIFAPGMTGASSAENNVDLMPROL"
complement(1703..1707)
/note="possible RBS for RV3494c"
complement(1706..2860)
/gene="lprN"
complement(1706..2860)
/gene="lprN"
complement(1706..2860)
/note="RV3495c" (MTV023.02c), len: 384. lprN, similar to
MTCY19H5.29 (402 aa) and (MTV051.08). Probably
lipoprotein, contains possible signal sequence and
appropriately positioned P500013 prokaryotic membrane
lipoprotein lipid attachment site. FASTA scores:
9p1297050|MTC128_13 (390 aa) opt: 653 z-score: 762.1 E():
0; 33.68 identity in 363 aa overlap; and
297182|MTCY19H5_29 (402 aa) opt: 572 z-score: 667.9 E():
1.1e-29; 31.88 identity in 362 aa overlap. Tbpase
score is 0.897"
/codon_start=1
/transl_table=11
/product="lprN"
/protein_id="CAI17732.1"
/db_xref="GI:2924432"
/db_xref="SPTREMBL:O53540"
/translation="MNRIMWLRAIILVASSALLAGCGFGLNSLPDGTAGHGCAVSV
TVMAVDVATLPQNSPYVVDVTVGSVAVVORPDSFYAAVKLIDLNNVLIPANV
AKVQSRLSGSLVHEVLAPEPTDPTGRLVDSRITRANTDREPTTEVSALGVVVK
AGVGALEIETDERHOAVAGROAFVNLVPELRLTAGLNQVDDIIDALDGLRVSAI
LADKDNIGALDITLPDAVYVLMONRDHIVDAFAAKRLTMVSHVLAETKVGEDL
KDIYSIVKALNDRKDFVYSIQLLLPFPNFEIKQAVKGDYINVTYTTDTLRRIGE
TFPTTAVFDPMNAHMDIILNPPDLIGELANLSGQAADPKIIPPGTASGO"
complement(2798..2830)
/gene="lprN"
/note="P500013 prokaryotic membrane lipoprotein lipid
attachment site"
complement(2857..4212)
/gene="RV3496c"
complement(2857..4212)
/gene="RV3496c"
/note="RV3496c" (MTV023.03c), len: 451. Unknown but
similar to Mycobacterium tuberculosis proteins MTC128.12
(530 aa), MTCY19H5.30c (508 aa) (MTV051.07). Hydrophobic
region at N-terminus. FASTA scores: 297050|MTC128_12 (530
aa) opt: 838 z-score: 711.0 E(): 4.3e-32; 35.18 identity
in 473 aa overlap; and 297182|MTCY19H5_30 (508 aa) opt:
821 z-score: 697.0 E(): 2.6e-31; 35.18 identity in 453 aa
overlap. Tbpase score is 0.891"
/codon_start=1
/transl_table=11
/product="hypothetical protein RV3496c"
/protein_id="CAI17733.1"
/db_xref="GI:2924433"
/db_xref="SPTREMBL:O53541"

VERSION	AP001081.1	GI:6815077
SOURCE	Oryza sativa (cultivar: Nipponbare) DNA, clone: P0693B08.	
ORGANISM	Oryza sativa Eukaryote; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.	
REFERENCE	1 (bases 1 to 140812) Sasaki, T., Matsumoto, T. and Yamamoto, K. Oryza sativa nippobare(GA3) genomic DNA, chromosome 1, PAC clone: P0693B08 Published Only in DataBase (2000) In press	
JOURNAL	2 (bases 1 to 140812) Sasaki, T., Matsumoto, T. and Yamamoto, K. Direct Submission Submitted (24-JAN-2000) to the DDBJ/EMBL/Genbank databases. Takui I Sasaki, National Institute of Agrobiological Resources, Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan (E-mail: tsasaki@agr.affrc.go.jp, URL: http://rpg.dna.affrc.go.jp/, Tel: 81-298-38-7441, Fax: 81-298-38-7468)	
AUTHORS	The orientation of the sequence is from SP6 to T7 of the PAC clone. Genes were predicted from the integrated results of the following: GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as Slicepredictor (October1998 version). The genomic sequence was searched against the non-redundant database NRP/PIR, SWISSPROT, GENEPT, PDB) from MAF DNABank and the cDNA sequence database at RCP. Protein similarities of the coding regions were searched against NRP with BLASTP2.0. ESTs represent the identified cDNA sequences using BLASTN2.0 with the corresponding DDBJ accession no. and RGP clone ID	
COMMENT	The sequence of this clone has an overlap with P0667A10 clone, DDBJ: AP001073 at the 5' end and an overlap with P0424A08 clone, DDBJ: AP000837 at the 3' end. This clone starts at the position 106,268 of P0667A10 and ends at the position 3,914 of P0424A08. Detailed information on assemble quality together with annotation of this entry at http://www.dna.affrc.go.jp:82/genomicdata/Genome?i=nished.html .	
FEATURES	Location/Qualifiers	
Source	1..140812	
	/organism="Oryza sativa"	
	/cultivar="Nipponbare"	
	/db_xref="taxon:4530"	
	/chromosome="1"	
CDS	/clone="P0693B08"	
	join(3816..4050,4056..4279,4294..4770)	
	/note="hypothetical protein"	
	/codon_start=-1	
	/protein_id="BA90364.1"	
CDS	/db_xref="GI:6815078"	
	/translation="MPHFSSLSITLGGSAAGDHRGACAEAAAGSGAAGCRRRRM EERAGGGAEGAGDQWRKQRAAGREGDREAIIVGSGAEVAQAAGDEEPP KEAAGCERSSNLLLAAMRAVQKAMRAREPTHGCGGCAETRAAAPTRSGH RAPTRRRSPPIARPPSPGPPPAVAALPEPPPAIGSAVPARRRPLLTATLLC RLRLADVAHVCAAVTLDVDRSPRRPPPTAPRRARCAPAFSTPVSAGPRL DHRPPPPPPVSLAFTLRSPSPTSASRPSPHP"	
	join(9496..9723,10535..11281)	
	/note="ESTs D24550(R2151), D24265(R1609), A0031848(R2151) correspond to a region of the predicted gene. Similar to cationic peroxidase isozyme 40K precursor (D42065)"	
	/codon_start=-1	
	/protein_id="BA90365.1"	
CDS	/db_xref="GI:6815079"	
	/translation="MDLAWFAVAVVCGLVGGSGAGLETNGLAVNFYQKSPND SYRSYVMAOVANPALPGRLLRHFHDCFVCGDASILDNAGSEKTAGPNLSVGY EVIDAIKTOLEACPGVSCADIVALAADAVSYOFKASLMOVETGRDQPSIASMT GALSPGAGFTLLSGPANGLTDLVALSGAHTIGKACSSVTPTLYOGNTSIDP LLSAYAKALMSSCRNPSSSTIDIDVATPLFSDGYANLQKOGALASDALTLON AAAAQVADLTNIFKYAFSSMSMKMRGRIDVLTGSGNTRKCRSAS"	
	join(13647..13769,14188..14391,14428..14602,14993..15468)	
	/note="hypothetical protein"	
	/codon_start=-1	
	/protein_id="BA90366.1"	
CDS	/db_xref="GI:6815080"	
	/translation="MRVERADIPVAMHOPPLLPGSSSIATADYEMNKQYGRAS AALLGTRPGRGGGGSGGSGFADGEBEGRLMWRVSCVSDSGGSGSLQASRF FLAISPISAMYSVYSYTSRSGRAPLDSAGNPFTTSLHRTSIDLETSRQPREL VLVASPLSPIMRRDRRMWMAFAGAPRTGSKVROIIEOCAISPCITLRLALCAL OLAEDHGAAPRVPCIOPGCEGLLHPQVDVGRALHACRLVGLGPPVGAACPGGL PSSACLRRGKREKREKENGKREEDVGCADICNCGSHADSAAT"	
	join(16052..16064,16146..16319,16355..16935)	
	/note="hypothetical protein"	
	/codon_start=-1	
	/protein_id="BA90367.1"	
CDS	/db_xref="GI:6815081"	
	/translation="MGPTLGAIVEGERGCDRMRSGDGRGREGVAGGARVARKRG RRMGEIGLALAAAGEETVGRGEERKAMHGAAMWERSGVAPTSLEFRLTNLACIP ALAAGLIALAGALPISFPSPRASSPKRRSSTLEPSCRSVAATADHNDCLARTKR EGYTGTGIRWRSSEPAADIDLTSTPLPRRCVRCFSPRLPRAQGRARTGIPASP SPKTCPTPTROPPLPSLAHRLHPLRTATPPAAGLSH"	
	18583..20061	
	/note="ESTs C71763(E0309), A0082059(E0309) correspond to a region of the predicted gene. Similar to Arabidopsis thaliana chromosome II BAC T20B5 genomic sequence: hypothetical protein. (AC002409)"	
	/codon_start=-1	
	/protein_id="BA90368.1"	
CDS	/db_xref="GI:6815082"	
	/translation="MAIIVGHCRLSPGCGVGRGRCRLRSVTQLDCPYFDHDLDD RRRNRHAGISPRSLSPGFTVVPADRAAPADDEKMPRVEIYAGGARVH ELIAEAGAATATCTRLVPAQGGIGGALLLEDGRSHVAVIKFLDADASSPSHGSGP YASKAVILEVAATLDDHGFARVEPALIKISRPAPTTTASIORFAHCDGCGGP SRSVASVHRIGSLDVRNLINDHAGNIIKVSSESGSGSTLPLDPLVDHILC LPRDLDDPYREIHMPOSSLPEGSALEVAVSDPRDAMLRAPLSITFEARILIT LCTIFORANAAGLCIADIGDMTRFESAMEBLSLSELCAKYATSSTTTASSSRK HHNSDSDSDSTFGMDVDPAGLPPLFLDGGGGLKSVSFAEGGAAGCAAKRM SFKALSDEMAALDPRFDLPAALDAKRRAGIKLRLGTSF"	
	complement(20793..21346)	
	/note="3' LTR"	
	join(21752..21851,22552..22676,22886..23101)	
LTR	/note="hypothetical protein"	
	/codon_start=-1	
	/protein_id="BA90369.1"	
	/db_xref="GI:6815083"	
	/translation="MDLKSTSTSVSLKLEQEGALPGRTMEREGGLNSETVGRI GOYMGSPITASNVVPLCEMNAEREAEALNRSRHTRAHATRRHATAAATAAP AASRSRRKSRAPPCGPPAMKOTSTASYAANGRQAGARDAS"	
CDS	complement(join(23394..23529,23558..24018))	
	/note="hypothetical protein"	
	/codon_start=-1	
	/protein_id="BA90370.1"	
	/db_xref="GI:6815084"	
CDS	/translation="KGMFARRTRRRRRDLAORLTAAGSGRSGCGGCKEDGAAV ECTGADCKTKRKEEGTGLVLTGSLVAGVAISPAEWRVGRERRERDSTKESRP RARTRAEERMAERQAQAATVATWPERGSEGGCQMRRLTEVDGDPDREEGGR LGWAFVGFMSAEGERRKMRPPIRRKRREKKIKRRK"	
	complement(join(24161..24471,24526..24637))	
	/note="hypothetical protein"	
	/codon_start=-1	
	/protein_id="BA90371.1"	
CDS	/db_xref="GI:6815085"	
	/translation="MAVTRVAVDEDEGPTGEGAVHHVALTWRDVAHTWQPHRTAAGS ARLARTGSSSRPGSGNVRRYRGAALADGSEGRERRRRDATTTRRRYRRKEGEE GGGVLEEVAPARRETTVNGETGROREVRKRRGSGTT"	
	join(25452..25542,26685..26854,26925..26983,27057..27489,27514..27825)	
	/note="EST A0082058(C12976) corresponds to a region of the predicted gene. Similar to a region of the hypothetical protein"	
	/codon_start=-1	
	/protein_id="BA90372.1"	
CDS	/db_xref="GI:6815086"	
	/translation="MERGGPYDCLCFRRIHLGTRGDCPLPAGDVGPLVDHQAASLX EDVAKASDAADVAATTSQGVNVPKKGKRSYVGNRRKPTPSPASDPPPRORLY TIGERAKARAAADDRSGNSASPVAASDVVYVNTDVPRTGPGDILVPOGSPAD IILTWELQYEMKRITLQAGRGITRHTIABARAAAASANERADKLADLAEVDEDLKM	

RELAVAGNEROGLKRNKSELENNLSIRGLSLRLARECGVTTTIPANPDEFLTITSLA
ELATAMEVIPSKEAARIGEETSSGIYTGTCNVLAACYKLAHPELDLQKILVQGAASNAR
KDYMERIGDGESEILPLFEER"
CDS complement(join(29314..30354,33082..35175))
/note="Similar to GAG-POL precursor (AB030283)
internal stop
codon:complement(29869..29871,30130..30132,30295..30297,
34924..34926,34960..34962)
probably inactive because stop codons are included in CDS"
/codon_start=1
/pseudo
LRR complement(30355..30714)
/note="3' LTR"
CDS complement(30656..31903)
/note="Similar to Zea mays 22-kDa alpha zein gene cluster;
hypothetical protein. (AF031569)"
/codon_start=1
/protein_id="BAA90373.1"
/db_xref="GI:6815087"
/translation="MAEPAKAHDLSPMSGDGEPPNRRARRPPPPRQSPRRERAL
ERAEGSATSTSTGGGGRGRRERLLVYGGSTPPGALQAAAGALLRPPPAHPESPA
QMLDVAKLVTAROLDAGRSATKASGAATGASARRARRAAAVRISATTP
SSTPSTREDLGGPDARTSLERRNGRRATHAEGSSSVQHGRRNPQSVPVGG
VGCRAVASLRNVRMPRRPRTIAEKYDGVNPAEFLQVYTTGIEAAGGDRYMANF
PMALGQARGLMNLNPPASVSWEDLCQFTMKFCGYPPGGEADLHAQRDDESL
RSYIORECOVRNTIPCIPTHAIVYAFRGVGRHNRMLEKIASKEPQTALFQLADRA
BMRKMGSPQKRRKCMWYPIKT"
LRR complement(31991..32353)
/note="5' LTR"
LRR complement(38386..38937)
/note="5' LTR"
join(40089..40184,40259..40396,40823..41078,41886..41957,
42532..42671)
/note="ESTs C91695(E31262), D40183(S1971) correspond to a
region of the predicted gene.
hypothetical protein"
/codon_start=1
/protein_id="BAA90374.1"
/db_xref="GI:6815088"
/translation="MAASAPPPAPPPPPPPAAAAATNPAPTAPTEDEGEVGCATPP
RTSALGTAMGTATLVGVFAGLLYGSKESASVSADAEVLMKSTTDRQYRLMR
DAMEKRPRIYAKSLVGLGMEFTAFEGIOMLLEKRGHDVFNAGSGATAAF
GLICSPMMRRARNTLVGALGAGVCFPLGWIQLKLAERANLEMASSKPTDLVEGQNS
GVGAATERLGRSLKK"
CDS complement(join(43206..43424,44141..44605,44699..45774,
43861..46014,46110..46206,46565..46701))
/note="Similar to glutamyl-tRNA synthetase (AF067773)"
/codon_start=1
/protein_id="BAA90375.1"
/db_xref="GI:6815089"
/translation="MEPKLAFQDSDPPLALICAAKAVGSLTIDPKLASGAPTLLHG

Db 72240 GGACGGCGG 72232
|| | ||
RESULT 10
HSE4
LOCUS
DEFINITION
HSE4
5880 bp DNA
Equine herpesvirus 4 (EHV4) genome, thymidine kinase (TK) and
glycoprotein H (gH) genes.
DI4486 D00683 D00684
DI4486.1 GI:221820
VERSION
KEYWORDS
UL21 gene; UL24 gene; gH gene; glycoprotein; thymidine kinase.
Equine herpesvirus 4 (strain 1942).
Equine herpesvirus 4
ORGANISM
Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
Alphan herpesvirinae; Varicelloviruses.
REFERENCE
AUTHORS
TITLE
1 (sites)
Nicolson,L., Cullinane,A.A. and Onions,D.E.
The nucleotide sequence of an equine herpesvirus 4 gene homologue
of the herpes simplex virus 1 glycoprotein H gene
J. Gen. Virol. 71 (Pt 8), 1793-1800 (1990)
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
2 (sites)
Nicolson,L., Cullinane,A.A. and Onions,D.E.
The nucleotide sequence of the equine herpesvirus 4 thymidine
kinase gene
J. Gen. Virol. 71 (Pt 8), 1801-1805 (1990)
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
3 (bases 1 to 5880)
Nicolson,L.
Unpublished (1993)
Submitted (10-SEP-1990) to DDBJ by:
Lesley Nicolson
Dept. Veterinary Pathology, University of Glasgow
Vet School
Beaarden Roda, Glasgow G61 1QH, Scotland
UK.
Phone: 041-330-8855
Fax: 041-330-5733.
Location/Qualifiers
1. 5880
/organism="Equine herpesvirus 4"
/db_xref="taxon:10331"
complement(99..104)
/note="put. polyadenylation signal (UL24); putative"
complement(115..120)
/note="put. polyadenylation signal (UL24); putative"
complement(116..934)
/note="putative; ORF1 (UL24 homologue)"
/codon_start=1
/protein_id="BAA03377.1"
/db_xref="GI:221821"
/translation="MKRKQRLTARSRLRACIRCHNRFYNNAMVDLASAKKNGYGARL
APFSELVPAETLKTAMGVSLAEFVNIGORPDVCVGFHGSDAKGVCLILELTC
RFSKNNMTASKNLORKGMROLDSCRLAKTLDPGSGELLAPLVFVNRGMRYLR
VTRLSPOVYVSNAAVLSCTISRAEYSPISSESTRRCYTRTFNKAAPAKTTGSI
OPTQAKPAATTAAVASLFSATQANTTAAGTQPATISLANPLAVASLFAK"
861..865
OPTQAKPAATTAAVASLFSATQANTTAAGTQPATISLANPLAVASLFAK"
87
/note="put. TATA box (TK); putative"
87
949..2007
/partial
/note="ORF2; (TK)"
/codon_start=1
/protein_id="BAA03378.1"
/db_xref="GI:221822"
/translation="MAACVPPGAPPSASCTPPRRQVTVIRIYLDGVYIGKSTGRV
MASAAGSGPTLIFPEPMATYRKLTFEFDVLSGIYDQNRKQGNLANVDAALITTAIO
SRFTTPLLIDHRTCTLFGNSLQGRQPOLTLVFRHPVASCFFPAARYLLGDSM
CALMAVATLPRPDGNTIVTTLNWEHRLRLTRARLGEODITLILATLRNVYEM
VNTCHFIRSGVRWDGSELPTSCGAYKHRTQMDAFOEVRSELDITLFALEPKTDEL
LDDRGVILEVHAMALDMLKLNLANVAFSADLSGTTRQCAAVVESLPLMSSTLSDFD

[illegible][illegible]

[illegible]

	/map="3p"	
	/note="isolated from a spinocerebellar ataxia 7 (SCA7)	
	patient"	
	115..477	
	/note="similar to NotI jumping clone sequence with GenBank	
	Accession Number X95831"	
	repeat_region	
	127..285	
	/note="region of cag expansion"	
	/rpt_unit=cag	
BASE COUNT	99 a	164 c 172 g 42 t
ORIGIN		
Query Match	6.8%; Score 40.4; DB 88;	Length 477;
Best Local Similarity	53.9%;	Pred. No. 1.2e+02;
Matches	83; Conservative	0; Mismatches 71; Indels 0; Gaps 0;
OY	145	gcgacggcggaaccggagtgtgtgtccctcgggagcatgacggcgagtagcaatcattacc 204
Db	334	GtGGCGGCGGTGGCGGGTCTCCTCCCTGGGGTGCGGAAGCCGCCGTCGTCCT 275
OY	205	aatacgaacttcctgcgcgcgtcgtcacccctgatcaccgccctgtttcagcgccagaac 264
Db	274	GCTGCTGCTGCCTGCGTGCCTGCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 215
OY	265	gcagcgcgcctcctccgc 298
Db	214	GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 181
RESULT 14		
LOCUS	HSN801297/c	
DEFINITION	HSM801297 3932 bp mRNA	PRI 18-FEB-2000
ACCESSION	AL133030	
VERSION	AL133030.1	GI:6453431
KEYWORDS		
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
TITLE	1 (bases 1 to 3932)	
JOURNAL	Koehler,K., Beyer,A., Mewes,H.W., Gassenhuber,J. and Wiemann,S. Direct Submission Submitted (15-NOV-1999) MIPS, Am Klopferspitz 18a, D-82152 Martinsried, GERMANY	
COMMENT	Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by BMFZ (Biomedical Research Center at the Charité, Berlin/Germany) within the cdna sequencing consortium of the German Genome Project.	
FEATURES	This clone (DKFP34H177) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcezentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at http://www.mips.biochem.mpg.de/proj/cdna/. Location/Qualifiers	
SOURCE	1. 3932	
gene	/organism="Homo sapiens"	
CDS	/db_xref="taxon:9606"	
	/clone="DKFP34H177"	
	/clone_id="434 (synonym: htes3). Vector pSport1; host	
	DH10B; sites NotI + SalI"	
	/dev_stage="adult"	
	/tissue_type="testis"	
	1..3505	
	/gene="DKFP34H177"	
	<1..3505	
	/note="DKFP34H177"	
	/note="similarity to peripheral benzodiazepine receptor	
	interacting protein"	
	/codon_start=2	

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2001, 23:31:44 ; Search time 330.77 Seconds
(without alignments)
1127.592 Million cell updates/sec

Title: US-09-138-735-1_COPY_1232_1825

Perfect score: 594

Sequence: 1 cagracagcgaacgactt.....aaacgagctcgtgacgata 594

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N.Geneseq_0601:*

- 1: /cgnl_9/gcgdata/geneseq/geneseq/NA1980.DAT:*
- 2: /cgnl_9/gcgdata/geneseq/geneseq/NA1981.DAT:*
- 3: /cgnl_9/gcgdata/geneseq/geneseq/NA1982.DAT:*
- 4: /cgnl_9/gcgdata/geneseq/geneseq/NA1983.DAT:*
- 5: /cgnl_9/gcgdata/geneseq/geneseq/NA1984.DAT:*
- 6: /cgnl_9/gcgdata/geneseq/geneseq/NA1985.DAT:*
- 7: /cgnl_9/gcgdata/geneseq/geneseq/NA1986.DAT:*
- 8: /cgnl_9/gcgdata/geneseq/geneseq/NA1987.DAT:*
- 9: /cgnl_9/gcgdata/geneseq/geneseq/NA1988.DAT:*
- 10: /cgnl_9/gcgdata/geneseq/geneseq/NA1989.DAT:*
- 11: /cgnl_9/gcgdata/geneseq/geneseq/NA1990.DAT:*
- 12: /cgnl_9/gcgdata/geneseq/geneseq/NA1991.DAT:*
- 13: /cgnl_9/gcgdata/geneseq/geneseq/NA1992.DAT:*
- 14: /cgnl_9/gcgdata/geneseq/geneseq/NA1993.DAT:*
- 15: /cgnl_9/gcgdata/geneseq/geneseq/NA1995.DAT:*
- 16: /cgnl_9/gcgdata/geneseq/geneseq/NA1996.DAT:*
- 17: /cgnl_9/gcgdata/geneseq/geneseq/NA1997.DAT:*
- 18: /cgnl_9/gcgdata/geneseq/geneseq/NA1998.DAT:*
- 19: /cgnl_9/gcgdata/geneseq/geneseq/NA1999.DAT:*
- 20: /cgnl_9/gcgdata/geneseq/geneseq/NA2000.DAT:*
- 21: /cgnl_9/gcgdata/geneseq/geneseq/NA2001.DAT:*
- 22: /cgnl_9/gcgdata/geneseq/geneseq/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	594	100.0	3402	20	AAx84092	T. cruzi PTC40 cod
2	590.8	99.5	3402	17	AAx727310	Trypanosoma cruzi
3	40.4	6.8	477	21	AAx244307	Human SCAT genomic
c	39.8	6.7	567	21	AAx29550	HIV codon altered
c	4	6.6	2790	19	AAx18480	BOP1 cDNA. Mus sp
c	6	6.4	2943	17	AAx16480	SAB virus 9B glyco
7	38	6.4	1037	21	AAx59242	Exons E, C and A o
8	38	6.4	1159	21	AAx59240	An Ecort Fragment
9	38	6.4	1472	21	AAx59241	Exons D, C, B and
10	37.6	6.3	2115	20	AAx32263	Human dunce-like p
11	37.6	6.3	2115	21	AAx88198	PPDB39 human dunce

12	37.6	6.3	72750	21	AAx81468
13	37.6	6.3	349980	21	AAx21544
14	37.2	6.3	921	20	AAx33558
15	37.2	6.3	966	20	AAx33558
16	37.2	6.3	2169	20	AAx33554
17	36.8	6.2	3737	21	AAx64656
18	36.8	6.2	3776	21	AAx42903
19	36.6	6.2	203	19	AAx30271
20	36.6	6.2	203	19	AAx17226
21	36.6	6.2	913	17	AAx30253
22	36.6	6.2	913	17	AAx13034
23	36.6	6.2	913	17	AAx70199
24	36.6	6.2	913	18	AAx70041
25	36.6	6.2	913	18	AAx62610
26	36.6	6.2	913	18	AAx35545
27	36.6	6.2	1966	20	AAx61220
28	36.6	6.2	1984	17	AAx13030
29	36.6	6.2	1985	17	AAx70036
30	36.6	6.2	1985	18	AAx60262
31	36.6	6.2	3324	20	AAx60262
32	36.4	6.1	2415	18	AAx6757
33	36.4	6.1	2453	18	AAx6757
34	36.2	6.1	633	11	AAx03322
35	36.2	6.1	862	21	AAx09682
36	36.2	6.1	3201	21	AAx14955
37	36	6.1	795	19	AAx55830
38	36	6.1	799	19	AAx55831
39	36	6.1	1448	21	AAx21931
40	36	6.1	1925	20	AAx50924
41	36	6.1	1926	21	AAx50254
42	36	6.1	2580	21	AAx5454
43	36	6.1	5452	20	AAx90923
44	36	6.1	8705	20	AAx23778
45	36	6.1	9600	19	AAx21683

ALIGNMENTS

RESULT 1	
AAx84092	standard; cDNA; 3402 BP.
AAx84092	
27-AUG-1999	(first entry)
T. cruzi PTC40 coding sequence.	
PTC40: Tc40; infection; diagnosis; immune complex; antigenic determinant; therapy; antibody; ds.	
Trypanosoma cruzi.	
WO9929867-A1.	
17-JUN-1999.	
10-DEC-1998;	98WO-IB01987.
10-DEC-1997;	97US-0988242.
(INMR) BIO MERIEUX.	
Jolivet M, Lesenechal M, Mandrand B, Paranhos-Baccala G;	
WPI; 1999-394978/33.	
P-PSDB; AAY22124.	
New Trypanosoma cruzi antigen	
Claim 1; Page 52-56; 65pp; English.	

CC This sequence encodes the Trypanosoma cruzi Prc100t protein of the
 CC invention, and is designated Tc40. The Prc100 antigenic determinant is
 CC useful as a reagent for detection and/or monitoring of Trypanosoma cruzi
 CC infection from samples including blood serum or plasma, saliva, or
 CC tears, by contacting with the sample and detecting an immune complex. The
 CC Prc100 antigenic determinant, the vector, expression cassette, cell or
 CC antibody are useful for treatment or prevention (vaccine) of a
 CC Trypanosoma cruzi infection in a man or animal. Current Trypanosoma cruzi
 CC antigens are obtained from protein fractions of the noninfectious stage
 CC of the parasite, and these do not allow sufficient production of antigens
 CC for use in reliable serological diagnostic tests. The strain to strain
 CC polymorphism reduces reliability of the tests.

CC Sequence 3402 BP; 888 A; 821 C; 956 G; 737 T; 0 other;

Query Match 100.0%; Score 594; DB 20; Length 3402;
 Best Local Similarity 100.0%; Pred. No. 8.3e-146;
 Matches 594; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 caggtacagcgttaacggtcttctgctcaatcgtacagccgagctgctgctgctgct 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 1232 caggtacagcgttaacggtcttctgctcaatcgtacagccgagctgctgctgct 1291
 OY 61 gacatgctgattcgatgacgattctccatctccggtctcccgagggagaagacagcag 120
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 1292 gacatgctgattcgatgacgattctccatctccggtctcccgagggagaagacagcag 1351
 OY 121 ccagggccaaaacacatcgtatgtagcgagcgagcgaggggtgtgtgtccctcgagcact 180
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 1352 ccagggccaaaacacatcgtatgtagcgagcgagcgaggggtgtgtgtccctcgagcact 1411
 OY 181 gacggcgagtagtagcagtcataccaatacgaactctgcgcgtgctgcatccctgataca 240
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 1412 gacggcgagtagtagcagtcataccaatacgaactctgcgcgtgctgcatccctgataca 1471
 OY 241 ccccccgttccggcgagcccaagagcgagcgccctcccgcgagcgagctgctgag 300
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 1472 ccccccgttccggcgagcccaagagcgagcgccctcccgcgagcgagctgctgag 1531
 OY 301 ccgacgctggggagcagaatctcgtatctagtagtaacatcgctggaggaatgaatgcacc 360
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 1532 ccgacgctggggagcagaatctcgtatctagtagtaacatcgctggaggaatgaatgcacc 1591
 OY 361 caaaggagcgtctgtaagcactggagcgccgagcagcagaggtctacggcggtgacgtcc 420
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 1592 caaaggagcgtctgtaagcactggagcgccgagcagcagaggtctacggcggtgacgtcc 1651
 OY 421 acgactacggcccgagcagaacatctcgtacgagcagacatggcgagccttggagcgt 480
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 1652 acgactacggcccgagcagaacatctcgtacgagcagacatggcgagccttggagcgt 1711
 OY 481 ggaattggtgagcagtaatagtgtgtgccaagcgagcctgctctcccaacgagcgagga 540
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 1712 ggaattggtgagcagtaatagtgtgtgccaagcgagcctgctctcccaacgagcgagga 1771
 OY 541 ccaacacagaggaagaaagcgctccgagcagctggaacagagctccgtggcgata 594
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 1772 ccaacacagaggaagaaagcgctccgagcagctggaacagagctccgtggcgata 1825

RESULT 2

AAT27310
 ID AAT27310 standard; cDNA; 3402 BP.

AC AAT27310;
 XX

DT 26-NOV-1996 (first entry)
 XX

DE Trypanosoma cruzi epimastigotic Prc100t antigen gene.
 XX

KM Antigen: Trypanosoma cruzi; epimastigote; serum; Chagas disease; probe;
 KW Primer; PCR; polymerase chain reaction; amplification; antibody; ds.

XX Trypanosoma cruzi.
 OS
 XX Key Location/Qualifiers
 FH CDS 266..3013
 FT /tag= a
 FT /product= Prc100t epimastigotic antigen

XX FR2723589-A1.

XX 16-FEB-1996.

XX 12-AUG-1994; 94FR-0010132.

XX 12-AUG-1994; 94FR-0010132.

XX (INMR) BIO MERIEUX.

XX Jollivet M, Lesenechal M, Paranhos-Baccala G;

XX WPI; 1996-190287/20.

XX P-PSDB; AAR91615.

PT New nucleic acid encoding Trypanosoma cruzi epimastigotic antigen -
 PT useful for diagnosis, monitoring and therapy of Chagas disease

XX Claim 1; Page 24-26; 55pp; French.

This is the nucleotide sequence encoding a novel isolated antigenic
 CC protein from Trypanosoma cruzi epimastigotes, designated Prc100t.
 CC The clone Tc50 was isolated from a T. cruzi genomic expression library in
 CC lambda gt10, using a mixture of sera from patients with Chagas disease.
 CC Clone Tc50 contained an 594 bp insert corresp. to nucleotides 1232-1825
 CC of this sequence. The Tc50 sequence was subsequently used to probe a
 CC Southern blot of restriction enzyme digested T. cruzi DNA and also screen
 CC a lambda gt10 library to isolate a 1041 bp EcoRI fragment corresp. to
 CC nucleotides 1403-2443 of Prc100t. Primers (AAT27311-5) were synthesised
 CC based on the sequences of the 594 and 1041 bp fragments and used to
 CC amplify the Prc100t clone as 3 fragments from cDNA derived from mRNA
 CC purified from T. cruzi epimastigotes. The protein or antibodies raised
 CC against it can be used in the detection and monitoring of T. cruzi
 CC infection i.e. Chagas disease.

XX Sequence 3402 BP; 888 A; 818 C; 958 G; 737 T; 0 other;

Query Match 99.5%; Score 590.8; DB 17; Length 3402;
 Best Local Similarity 99.7%; Pred. No. 5.7e-145;
 Matches 592; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 caggtacagcgttaacggtcttctgctcaatcgtacagccgagctgctgctgctgct 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 1232 caggtacagcgttaacggtcttctgctcaatcgtacagccgagctgctgctgctgct 1291
 OY 61 gacatgctgattcgatgacgattctccatctccggtctcccgagggagaagacagcag 120
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 1292 gacatgctgattcgatgacgattctccatctccggtctcccgagggagaagacagcag 1351
 OY 121 ccagggccaaaacacatcgtatgtagcgagcgagcgaggggtgtgtgtccctcgagcact 180
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 1352 ccagggccaaaacacatcgtatgtagcgagcgagcgaggggtgtgtgtccctcgagcact 1411
 OY 181 gacggcgagtagtagcagtcataccaatacgaactctgcgcgtgctgcatccctgataca 240
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 1412 gacggcgagtagtagcagtcataccaatacgaactctgcgcgtgctgcatccctgataca 1471
 OY 241 ccccccgttccggcgagcccaagagcgagcgccctcccgcgagcgagctgctgag 300
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 1472 ccccccgttccggcgagcccaagagcgagcgccctcccgcgagcgagctgctgag 1531
 OY 301 ccgacgctggggagcagaatctcgtatctagtagtaacatcgctggaggaatgaatgcacc 360
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 1532 ccgacgctggggagcagaatctcgtatctagtagtaacatcgctggaggaatgaatgcacc 1591

CC the development of serological immunoassays. One approach is to
CC synthesize peptides which, based on the properties of the predicted
CC protein sequence, are likely to be immunologically active. Such
CC peptides can be used as substrate antigens in immunoassays to detect
CC serum antibodies which recognize this specific peptide sequence.
CC Synthetic peptides may also be used to produce antibodies against
CC specific regions of the gp glycoprotein which are unique to one
CC virus. These can then be used to develop virus-specific
CC immunoassays for differentiation of SAb virus from other primate
CC alpha-herpes viruses and for identification of antibodies directed
CC against SAb virus in primate serum samples.

XX Sequence 2943 BP; 462 A; 1103 C; 983 G; 395 T; 0 other;

Query Match 6.4%; Score 38.2; DB 17; Length 2943;
Best Local Similarity 47.7%; Pred. No. 0.94; Mismatches 123; Indels 0; Gaps 0;
Matches 112; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

QY 209 cgaattctgcgcgtgcgcatccctgcatacccccgtgttcagcgccagccaagcag 268
DB 255 CGTCTCTGTCGGGAGACCTCTCTGCGGAGACGCGGCGTGGCGGCGGCGCGG 196
QY 269 ccgcgcctctgc 328
DB 195 TCGGGGCG 136
QY 329 atctagtgatcagctggggttaattatcaccacaagagcgtgtgcacatcgagcgc 388
DB 135 GGAAGACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 76
QY 389 cgccacagcagaggtctacagcggtgacgtccacgactacgccccagcgaac 443
DB 75 CGGGCGCTACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 21

RESULT 7
ID AAA59242 standard; cDNA; 1037 BP.

XX AAA59242;

XX 07-NOV-2000 (first entry)

DE Exons E, C and A of the SCA8 repeat region.

KM Repeat region: spinocerebellar ataxia type 8; SCA8; long arm;
KM chromosome 13; polymorphic CTA repeat; CTG repeat; ss.

XX Homo sapiens.

PN CA2283758-A1.

PD 28-APR-2000.

PF 27-SEP-1999; 99CA-2283758.

PR 28-OCT-1998; 98US-0181585.

PA (MINU) UNIV MINNESOTA.

PI Benzow KA, Moseley-allredge ML, Ranum LPW, Koob MD;

XX WPI: 2000-491456/44.

PT Novel nucleic acid molecule containing repeat region of spinocerebellar
PT ataxia type 8 coding sequence, useful as primer or probe for diagnosing
PT or determining risk for developing spinocerebellar ataxia -

XX Claim 25; Fig 7C; 68bp; English.

XX The present sequence represents a nucleic acid molecule containing a
CC repeat region of a spinocerebellar ataxia type 8 (SCA8) coding

CC sequence which is located within the long arm of chromosome 13. The
CC SCA8 coding sequence contains polymorphic CTA repeats and CTG repeats.
CC Generally, the CTA repeat is unstable, and can vary between individuals
CC in different families (the number of CTA repeats in the region does not
CC vary between individuals of a family). Fragments of the nucleic acid
CC molecule are useful as primers and probes for determining whether an
CC individual has or is at risk of developing spinocerebellar ataxia 8.

XX Sequence 1037 BP; 229 A; 261 C; 242 G; 305 T; 0 other;

Query Match 6.4%; Score 38; DB 21; Length 1037;
Best Local Similarity 56.3%; Pred. No. 0.8; Mismatches 71; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 173 cgggacgcagc 232
DB 629 ctggcttactactactactactactactactactactactactactactactactctgctg 688
QY 223 ctgcatcacccctgttcaagcgccagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 292
DB 689 ctgc 748
QY 293 cggcctg 298
DB 749 ctgcctg 754

RESULT 8
ID AAA59240 standard; DNA; 1159 BP.

XX AAA59240;

XX 07-NOV-2000 (first entry)

DE An EcoRI fragment that includes SCA8 repeat region.

KM Repeat region: spinocerebellar ataxia type 8; SCA8; long arm;
KM chromosome 13; polymorphic CTA repeat; CTG repeat; ss.

XX Homo sapiens.

PN CA2283758-A1.

PD 28-APR-2000.

PF 27-SEP-1999; 99CA-2283758.

PR 28-OCT-1998; 98US-0181585.

PA (MINU) UNIV MINNESOTA.

PI Benzow KA, Moseley-allredge ML, Ranum LPW, Koob MD;

XX WPI: 2000-491456/44.

PT Novel nucleic acid molecule containing repeat region of spinocerebellar
PT ataxia type 8 coding sequence, useful as primer or probe for diagnosing
PT or determining risk for developing spinocerebellar ataxia -

XX Claim 26; Fig 7A; 68bp; English.

XX The present sequence represents a nucleic acid molecule containing a
CC repeat region of a spinocerebellar ataxia type 8 (SCA8) coding
CC sequence which is located within the long arm of chromosome 13. The
CC SCA8 coding sequence contains polymorphic CTA repeats and CTG repeats.
CC Generally, the CTA repeat is unstable, and can vary between individuals
CC in different families (the number of CTA repeats in the region does not
CC vary between individuals of a family). Fragments of the nucleic acid
CC molecule are useful as primers and probes for determining whether an
CC individual has or is at risk of developing spinocerebellar ataxia 8.

QY 173 cgggcactgacgcgagtagcagtcataccaatacgactctcgccgctgctgcacccc 232

Query match 6.3%; Score 37.6; DB 20; Length 2115;

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2001, 16:22:20 : Search time 4309.39 seconds
(without alignments)
1302.967 Million cell updates/sec

Title: US-09-138-735-1_COPY_1232_1825
Perfect score: 594
Sequence: 1 caggtaacgagctaacgagctt.....aaacgagctccgtggcgata 594

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1022815 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estp11:*
73: em_estp12:*
74: em_estp13:*
75: em_estp14:*
76: em_estp15:*
77: em_estp16:*
78: em_estp17:*
79: em_estp18:*
80: em_estp19:*
81: em_estp110:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est33:*
111: gb_est34:*
112: gb_est35:*
113: gb_est36:*
114: gb_est37:*
115: gb_est38:*
116: gb_est39:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

JOURNAL

Submitted (02-JUN-1999) Genoscope - Centre National de Séquençage
BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)

COMMENT

- Web : www.genoscope.cns.fr

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see [http://www.fruitfly.org/The BDGP Drosophila melanogaster BAC library](http://www.fruitfly.org/The%20BDGP/Drosophila%20melanogaster/BAC%20library.html) was prepared by Kazuo Oosegawa and Aaron Mammeter in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCL-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source

1..925

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone_lib="RPCL-98"

/clone="BACR19D16"

/note="end : TET3"

BASE COUNT 120 a 61 c 61 g 172 t 511 others

ORIGIN

Query Match 8.0%; Score 47.8; DB 219; Length 925;
Best Local Similarity 11.8%; Pred. No. 0.024;
Matches 37; Conservative 151; Mismatches 125; Indels 0; Gaps 0;

OY 256 ccagccaaggcgcgcgcgcctcccgccggcgccgatccgctgaagcgcacgtggggagc 315
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 914 SSBGCSSTSMSTSSNSBBCSCSSBSBSTSMSSBSBSSSSSTTSACVCKC 855
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 316 aaatcatcttgtaactagtatcaatcagcttgggattaatatgttcaccacaaggagcgtgc 375
||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::
Db 854 NASSCCGCCCGMABCCKCSSSCCGGSASARGVKVRAGRGGRGGSGASASHSSS 795
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::
OY 376 agcacatggagcgcgcgcgcacagaggtcctaagcgggtggaagctcacgactacgccccc 435
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::
Db 794 ACBSSSSSCSAWMASSSSSSASSSRSRSGGAGGSGASSSSSSSSASASGVSSAS 735
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::
OY 436 cagcgaacaagctaacgagcacatgtgcgcacctgtgcagcgtcgatttgtgcagct 495
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::
Db 734 SSSSSCSTSVGCSSVASMSCSBSSBSASASSSSSSSASACSCCTTSMSCSTSA 675
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::
OY 496 aatagttgctcagcgcgcgtctgtctccacagcgcgcgaaaccaagaagaga 555
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::
Db 674 SMSAARSSTSSSSSCSSSMSASASSSSASSSSSSSSSSGSACGBMSGGGSGSV 615
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::
OY 556 aagcctccgcgcg 568
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::
Db 614 SASGCMSSSVSS 602

RESULT 2

AM922844

LOCUS AM922844 466 bp mRNA EST 19-JUL-2000

DEFINITION DG1_46_C04.g1_A002 Dark Grown 1 (DG1) Sorghum bicolor cDNA, mRNA sequence.

ACCESSION AM922844

VERSION AM922844.1 GI:8088669

KEYWORDS EST.

SOURCE Sorghum.

ORGANISM Sorghum bicolor

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliopsida; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE 1 (bases 1 to 466)
Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt

TITLE
JOURNAL
COMMENT

L.H.
An EST database from Sorghum: dark-grown seedlings
Unpublished (2000)
Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmp@prattuga.edu
Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: PolyTMix
High quality sequence start: 135
High quality sequence stop: 466
POLYA-No.

FEATURES
source

Location/Qualifiers
1..466
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Dark Grown 1 (DGI)"
/note="Organ: 5-day-old dark-grown seedlings; Vector:
Lambda Zap; Site_1: XhoI; Site_2: EcoRI; The library was
made from poly-A RNA in the cloning vector Lambda Zap II.
Clones to be sequenced were prepared by mass excision."
BASE COUNT 67 a 121 c 182 g 95 t 1 others
ORIGIN

Query Match 7.4%; Score 44; DB 122; Length 466;
Best Local Similarity 49.8%; Pred. No. 0.22;
Matches 110; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

QY 210 gactctcgcgtctgtcgtatccctcgtatccctcgtttcagcgcaagcagcgc 269
DB 30 gctacgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 89
QY 270 cgcgcctcgc 329
DB 90 ggc 149
QY 330 tctagtgaatcagctggggaatgaatgaatgaatgaatgaatgaatgaatgaat 389
DB 150 gctgctcgc 209
QY 390 ggcacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 430
DB 210 ggcgcgtacggtgacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 250

RESULT 3
CNS010BS 1036 bp DNA GSS 26-JUL-1999
LOCUS
DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC
BACN03H11 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL098770
VERSION AL098770.1 GI:5610381
KEYWORDS GSS.
SOURCE
ORGANISM
fruit fly.
Plasmod Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 1036)
Genoscope.
Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seque@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -

http://www.edgp.ebi.ac.uk - : This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billand at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelOBAC11.

FEATURES
source

Location/Qualifiers
1..1036
/organism="Drosophila melanogaster"
/plasmid="pBelOBAC11"
/db_xref="taxon:7227"
/clone_lib="DrosBAC"
/clone="BACN03H11"
/note="end : SP6"
BASE COUNT 240 a 127 c 155 g 234 t 280 others
ORIGIN

Query Match 7.4%; Score 43.8; DB 219; Length 1036;
Best Local Similarity 23.6%; Pred. No. 0.27;
Matches 92; Conservative 113; Mismatches 185; Indels 0; Gaps 0;

QY 94 cgtctcccgaggaagaaacagcagcagcagcagcagcagcagcagcagcagcagc 153
DB 637 CAACCCSSAASVMSGASARAAAVSAVAGMGAAGAAAGAAAGAAAGAAAGAAAG 696
QY 154 aaacgggagtgctgtcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 213
DB 697 CVCVCYKCGCCGCCYGAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 756
QY 214 tctgctcgc 273
DB 757 RAVCGCCG 816
QY 274 cctctcgc 333
DB 817 SCSCGASCCSVSSGCG 876
QY 334 gtaatcagctgtggaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 393
DB 877 SGTSGGSSGSSGCGAG 936
QY 394 acgacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 453
DB 937 ASMGCKGMAAAASVSSGCGCAAAAAAARASGCGSSGCGSSGCGSSGCGSSGCG 996
QY 454 gggcaaatgacgcagcagcagcagcagcagcagcagcagcagcagcagcagc 483
DB 997 GSGCGSAGCGSSGCGSSGCGSSGCGSSGCGSSGCGSSGCGSSGCGSSGCGSSG 1026

RESULT 4
AM677864 558 bp mRNA EST 19-JUL-2000
LOCUS
DEFINITION WSI_11_G05_g1_A002 Water-stressed 1 (WS1) sorghum bicolor cDNA,
mRNA sequence.
ACCESSION AM677864
VERSION AM677864.1 GI:7551577
KEYWORDS EST.
SOURCE
ORGANISM
sorghum.
Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE
1 (bases 1 to 558)
Cordonnier-Pratt M.-M., Gingle A., Marsala C., Sudman M. and Pratt
L.H.
An EST database from Sorghum: water-stressed plants
Unpublished (2000)
Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Query Match	7.0%;	Score 41.4;	DB 247;	Length 559;
Best Local Similarity	58.5%;	Pred. No. 1.1;		
Matches	72;	Conservative	0;	Mismatches 51;
			Indels	0;
			Gaps	0;

Oy 165 tgtgtcctcggaactgacgcggagtagcaatcataccaatacgaattctgcagtc 224
|| || || || || || || || || || || || || || || || || || || ||
Db 86 TGGCTGTGCGCCCGTGCATGCCCGTGCTCCTGCTGTGCTGCGCGCTGTC 145

Qy 225 tgcattccctgcatcacccctgttttaagcgagccaaagcagccgcctctccgc 284
 ||| ||||| | ||||| | | ||||| ||| | ||||| |
 Db 146 TGCATCCCTGCTGCTGCCCTGCTGCTGCTGCCAATGCTGCCGATGCTGCTGCC 205

QY	285	ggc	287
Db	206	TGC	208

RESULT	13
AA858891/c	
LOCUS	AA858891
DEFINITION	447 bp mRNA EST 03-JUL-1999 UT-R-A0-bd-d-01-0-UT.s1 UT-R-A0 Rattus norvegicus cDNA clone

ACCESSION	AA858891	
VERSION	AA858891.1	GI:4223084
KEYWORDS	EST.	
SOURCE	Norway rat.	
ORGANISM	Rattus norvegicus	

ORGANISM *Rattus norvegicus*
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE	1 (bases 1 to 447)
AUTHORS	Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE	Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL	Genome Res. 6 (9), 791-806 (1996)
MEDLINE	97044477
COMMENT	On Mar 10, 1998 this sequence version replaced gi:2948242.

Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565

Oligo-dt track not found. Not I site shown in beginning of sequence is likely internal to the message. cDNA library Preparation: M. Fatima Bonaldo, Ph.D. Clone distribution: clones will be available through Research Genetics This clone is also available through the I.M.A.G.E. Consortium at LNL (infoimage.llnl.gov). IMAGE ID=176747 The following repetitive elements were found in this cDNA sequence: 15-121, >(CAG)n#simple_repeat 61-179, >(CGC)n#simple_repeat
Seq primer: M13 Forward
POLYA-NO.

FEATURES	Location/Qualifiers
source	1. .447

```

1. .447
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db.xref="taxon:10116"
/clone="UI-R-A0-bd-d-01-0-01"
/clone_lib="UI-R-A0"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: p773D-pac (Pharmacia) with a modified
polylinker. Site_1: Not I. Site_2: Eco RI. This library
consists of a mixture of individually tagged normalized
libraries constructed from rat placenta, adult lung, brain
liver, kidney, heart, spleen, ovary, and muscle. The tag
is a string of 3-5 nucleotides present between the Not I
site and the oligo-dT track which allows identification of
the library of origin of a clone within the mixture."

```

BASE COUNT	125 a	150 c	110 g	62 t
ORIGIN				

Query Match	6.98;	Score 41.2;	DB 12;	Length 447;
Best Local Similarity	52.38;	Pred. No. 1.2;		
Matches 91; Conservative	0;	Mismatches 83;	Indels 0;	Gaps 0;

OY 137 cggtagtgccgacgcgcgaaccggggtgtgtctcctcggcactaacgcgcagatagca 196
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 177 cggcggcggaaggcgcgctgcccgcgccgcggcgctgcttcgccaaggcgccgccccgttg 118

Qy 197 gtcataccaatagcacttctgcgcgctcgtcatccctcgataccccctgtttcacgyc 256
| | | | | | | | | | | | | | | |
Db 117 ctgcctgcctgcctgcctgcctggggcgctggcgggcgctgctgctgctgctggggcgcttg 58

Oy 257 cagccaaagcagccgcgctcctcctccgcgcgcgatcgctgagccgaactlg 310
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 57 ccggcggcggcttgcgggggctgtgtgtttgtctgtcacccggggcggg 4

RESULT 14				
CNS0091P				
LOCUS	CNS0091P	925 bp	DNA	GSS
DEFINITION	Drosophila melanogaster genome survey sequence TET3 end of BAC			

ACCESSION	AL053013
VERSION	AL053013.1
KEYWORDS	GS.
SOURCE	fruit fly.
ORGANISM	Drosophila melanogaster

REFERENCE
AUTHORS
TITLE
JOURNAL

Eukaryota: Metazoa: Arthropoda: Tracheata: Hexapoda: Insecta: Psycoropta: Neoptera: Endopterygota: Diptera: Brachycera, Muscomorpha, Ephydroidea: Drosophilidae, Drosophila.
1 (bases 1 to 925)
.Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage

COMMENT

Determination of this BAC-end sequence was carried out as part of a Collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the *Drosophila melanogaster* genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP *Drosophila melanogaster* BAC library was prepared by Kazuhiro Oseegawa and Aaron Mammosser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPci-98 and was constructed by partial EcoRI digestion of *Drosophila* DNA provided by the BDGP from the isogenic strain Y2: cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES	Location/Qualifiers
source	1. .925

	BASE COUNT	ORIGIN
/oranism="dirosophila melanogaster"		
/db.xref="taxon:7227"		
/clone.lib="RRC1-98"		
/clone="BACRI9D16"		
/note="end : TET3"		
120 a	61 c	172 t
	61 g	511 others

[illegible]

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2001, 18:26:07 : Search time 164.23 Seconds
(without alignments)
684.715 Million cell updates/sec

Title: US-09-138-735-1_COPY_1232_1825

Perfect score: 594

Sequence: 1 caggtacgcgcgtacgcgtc.....aaacgcgtcctgctgcgata 594

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 94655562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA:*

1: /cgnl_7/prodata/1/ina/5A.COMB.seq:*
2: /cgnl_7/prodata/1/ina/5B.COMB.seq:*
3: /cgnl_7/prodata/1/ina/6A.COMB.seq:*
4: /cgnl_7/prodata/1/ina/6B.COMB.seq:*
5: /cgnl_7/prodata/1/ina/PCRTUS.COMB.seq:*
6: /cgnl_7/prodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	594	100.0	3402	1	US-08-480-917-1
2	39.2	6.6	2790	2	US-08-718-661-1
3	38.2	6.4	2943	1	US-08-042-747A-7
4	37.6	6.3	2115	2	US-08-474-379C-60
5	37.6	6.3	2115	3	US-09-146-249A-60
6	37.6	6.3	2115	3	US-08-206-188B-60
7	36.8	6.2	3777	3	US-09-121-321-15
8	36.8	6.2	3777	4	US-08-933-803A-15
9	36.6	6.2	2033	4	US-09-043-303-7
10	36.6	6.2	913	1	US-08-217-327-3
11	36.6	6.2	913	1	US-07-885-970A-3
12	36.6	6.2	913	1	US-08-298-687A-3
13	36.6	6.2	913	1	US-08-530-797-2
14	36.6	6.2	913	1	US-08-298-829-3
15	36.6	6.2	913	2	US-08-787-335-2
16	36.6	6.2	1984	1	US-07-885-970A-25
17	36.6	6.2	1985	1	US-08-298-687A-25
18	36.6	6.2	1985	1	US-08-298-829-25
19	36.6	6.1	2580	3	US-09-050-863-2
20	36.6	6.1	5452	2	US-09-130-114-1
21	36.6	6.1	9600	4	US-08-910-647-1
22	36.6	6.1	10596	1	US-07-884-811-15
23	36.6	6.1	10596	1	US-07-885-971-15
24	36.6	6.1	10596	1	US-08-087-783A-15
25	36.6	6.1	10596	1	US-08-194-088B-15
26	36.6	6.1	10596	2	US-08-194-087-15
27	36.6	6.1	10596	5	PCT-US93-04648-15

C 28	35.2	5.9	3376	1	US-08-320-559-29	Sequence 29, Appl
C 29	35.2	5.9	3376	3	US-08-545-860D-29	Sequence 29, Appl
C 30	35.2	5.9	3376	5	PCT-US94-00496-29	Sequence 29, Appl
C 31	35.2	5.9	168	1	US-08-469-802B-4	Sequence 4, Appl
C 32	35.2	5.9	168	2	US-08-267-803B-4	Sequence 4, Appl
C 33	35.2	5.9	171	1	US-08-469-802B-5	Sequence 5, Appl
C 34	35.2	5.9	171	2	US-08-267-803B-5	Sequence 5, Appl
C 35	35.2	5.9	195	1	US-08-469-802B-2	Sequence 2, Appl
C 36	35.2	5.9	195	2	US-08-267-803B-2	Sequence 2, Appl
C 37	35.2	5.9	234	1	US-08-469-802B-3	Sequence 3, Appl
C 38	35.2	5.9	234	2	US-08-267-803B-3	Sequence 3, Appl
C 39	35.2	5.9	336	2	US-07-814-220-3	Sequence 3, Appl
C 40	35.2	5.9	336	2	US-07-812-421-3	Sequence 3, Appl
C 41	34.8	5.9	1299	1	US-07-661-610C-11	Sequence 11, Appl
C 42	34.8	5.9	2001	1	US-08-674-168-24	Sequence 24, Appl
C 43	34.8	5.9	2001	3	US-08-985-908-10	Sequence 10, Appl
C 44	34.8	5.9	2001	3	US-08-852-730-23	Sequence 23, Appl
C 45	34.8	5.9	5437	1	US-07-661-610C-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-480-917-1
: Sequence 1, Application US/08480917
: Patent No. 5820864
: GENERAL INFORMATION:
: APPLICANT: PARANHOS-BACCALA, Glaucia
: APPLICANT: LESENECHAL, Mylene
: APPLICANT: JOLIVET, Michel
: TITLE OF INVENTION: NEW TRYANOSOMA CRUZI ANTIGEN, AND GENE
: TITLE OF INVENTION: ENCODING THE LATENT: THEIR APPLICATION TO THE DETECTION OF
: NUMBER OF SEQUENCES: 13
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Oliff & Berridge
: STREET: 700 South Washington Street, Suite 300
: CITY: Alexandria
: STATE: Virginia
: COUNTRY: U.S.A.
: ZIP: 22314
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/480,917
: FILING DATE: 07-JUN-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Berridge, William P.
: REGISTRATION NUMBER: 30,024
: REFERENCE/DOCKET NUMBER: WPB 36400
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 703-836-6400
: TELEFAX: 703-836-2787
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3402 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: US-08-480-917-1
Query Match 100.0%; Score 594; DB 1; Length 3402;
Best Local Similarity 100.0%; Pred. No. 1; 1e-152;
Matches 594; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
0y 1 caggtacgcgcgtacgcgttcctcaatcgtacagcgcgtacgtacgtcctcgtcgt 60

[illegible]

```

RESULT 2
US-08-718-661-1
; Sequence 1, Application US/08718661
; Patent No. 5876972
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Nucleic acid molecules coding for mammalian
; TITLE OF INVENTION: tumor suppressor proteins and methods for their isolation
; NUMBER OF SEQUENCES: 15
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/718,661
; FILING DATE:
; CLASSIFICATION: 530
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2790 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 542..2545
; US-08-718-661-1

```

Query Match	6.68;	Score 39.2;	DB 2;	Length 2790;
Best Local Similarity	54.18;	Pred. NO. 0.093;		

	Matches	80;	Conservative	0;	Mismatches	68;	Indels	0;	Gaps	0;
OY	137	cggtagtcgacgaagcgaaacccggagtgatgtgtctctcgagcaatcagcgcggcgaatgaaga	196							
Db	2114	CAGCGACGACGCCACCCGCACTGGCGGCCACGACCGCTCCCACTGGCCACAGGCACACTGCCA	2173							
OY	197	gfcataccaatcagactctgcgcgtctgatactccctcatcacccccctgtttcaagcgc	256							
Db	2174	GGCGCACAGCCACTACCAAGCCACAGATGACAGCCACAGTTTCACTTGCAGATCCAGCCC	2233							
OY	257	caagcaagcagcgcgcgcctcctgcgcgc	284							
Db	2234	CAGATGCAAGCTTACCAACAGCTGCTGCCGC	2261							

```

1      RESULT      3
2      US-08-042-747A-7/c
3      ; Sequence 7, Application US/08042747A
4      ; Patent No. 5487969
5      ;
6      GENERAL INFORMATION:
7      APPLICANT: Eberle, Richard
8      APPLICANT: Black, Darla
9      APPLICANT: Scinicariello, Franco
10     APPLICANT: Hilliard, Julia K.
11     TITLE OF INVENTION: Cloning and Amplification of Monkey B
12     TITLE OF INVENTION: Virus Genes
13     NUMBER OF SEQUENCES: 12
14     CORRESPONDENCE ADDRESSES:
15     ADDRESSEE: Cox & Smith Incorporated
16     STREET: 112 East Pecan Street, Suite 2000
17     CITY: San Antonio
18     STATE: Texas
19     COUNTRY: USA
20     ZIP: 78205
21     COMPUTER READABLE FORM:
22     MEDIUM TYPE: Floppy disk
23     COMPUTER: IBM PC compatible
24     OPERATING SYSTEM: PC-DOS/MS-DOS
25     SOFTWARE: PatentIn Release #1.0, Version #1.25
26     CURRENT APPLICATION DATA:
27     APPLICATION NUMBER: US/08/042,747A
28     FILING DATE:
29     CLASSIFICATION: 435
30     ATTORNEY/AGENT INFORMATION:
31     NAME: Haymond, W. Bradley
32     REGISTRATION NUMBER: 35186
33     REFERENCE/DOCKET NUMBER: S-0072.179
34     TELECOMMUNICATION INFORMATION:
35     TELEPHONE: 210-554-5500
36     TELEFAX: 210-226-8395
37     TELEX: 767609
38     INFORMATION FOR SEQ ID NO: 7:
39     SEQUENCE CHARACTERISTICS:
40     LENGTH: 2943 base pairs
41     TYPE: nucleic acid
42     STRANDEDNESS: single
43     TOPOLOGY: linear
44     MOLECULE TYPE: Genomic DNA
45     FEATURE:
46     NAME/KEY: CDS
47     LOCATION: 87..2744
48     US-08-042-747A-7

```

	Query Match Similarity	6.4%	Score 38.2	DB 1	Length 2943
	Best Local Similarity	47.7%	Pred No. 0.18		
	Matches	112	Conservative	0	Mismatches 123; Indels 0; Gaps 0
OY	cgactctgcgcgtcgtgatcccttcgataccccccttgatttcagggccaggcaagacag	268			
Dd	ccttctgcgtccggaccttcctcgtcctccggaaaccggccgccggcgatccggc	196			
OY	cgcgcctctctgcgcgcgcgcgatcgcgtctgaacgcgcacgttgaggacaagaatcatgcta	328			

Db 195 TCGGGGCGCCGCCCGCGCGGGGGGGGTCGTCGGGCGGGGGCATCACGGCGAGGGCGA 136
 Oy 329 atctagtgaaatcagctgaggatlaatgtcaccaagaagacgltcgtacgactggagcgc 388
 Db 135 GGAGACGGGGGAGGGGGAAGAAAGAGGGGGGGTGTCCGCGAGGCCGATCCGGGCGCG 76
 Oy 389 cggcgcagcagcaggtctcagcggttgacgtcccaagataccgccgcgcagcgaac 443
 Db 75 CGGGGCTACCCGGGCGACACTACCGGGGGGTGTGATCTCGGGCGCGGGCGGGGAC 21

```

RESULT 4
US-08-474-379C-60
Sequence 60 Application US/08474379C
Patent No. 5977305
GENERAL INFORMATION:
APPLICANT: Wigler, Michael H.
APPLICANT: Colicelli, John J.
TITLE OF INVENTION: CLONING BY COMPLEMENTATION AND RELATED
TITLE OF INVENTION: PROCESSES
NUMBER OF SEQUENCES: 88
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive/6300 Sears Tower
City: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/474,379C
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/511,715
FILING DATE: 20-APR-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/206,188
FILING DATE: 01-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/688,352
FILING DATE: 19-APR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 27866/32771
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
INFORMATION FOR SEQ ID NO: 60:
SEQUENCE CHARACTERISTICS:
LENGTH: 2115 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna
FEATURE:
NAME/KEY: CDS
LOCATION: 762..2115
US-08-474-379C-60

```

Query Match	6.3%	Score 37.6	DB 2	Length 2115
Best Local Similarity	63.0%	Pred. No. 0.23		
Matches 58	Conservative 0	Mismatches 34	Indels 0	Gaps 0
07	212	cttcgcgcgtctgcataccctgcataccccctggtttcagcgcagcaagcagccg	271	

Db 577 CTGGCCCCCGCGCGCGCCACCCCTGCGCCCGACGCCCCCAGTGTACGCGCGCGCGCGCG 636

```

: RESULT 5
: US-09-146-249A-60
: Sequence 60, Application US/09146249A
: Patent No. 6069240
: GENERAL INFORMATION:
: APPLICANT: Wigler, Michael H.
: APPLICANT: Colicelli, John J.
: TITLE OF INVENTION: Cloning by Complementation and Related
: TITLE OF INVENTION: Processes
: NUMBER OF SEQUENCES: 85
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
: STREET: 6300 Sears Tower, 233 South Wacker Drive
: CITY: Chicago
: STATE: Illinois
: COUNTRY: United States of America
: ZIP: 60606-6402
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/146,249A
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/511,715
: FILING DATE: 20-APR-1990
: ATTORNEY/AGENT INFORMATION:
: NAME: Clough, David W.
: REGISTRATION NUMBER: 36,107
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 312/474-6300
: TELEFAX: 312-474-0448
: TELEX: 25-3856
: INFORMATION FOR SEQ ID NO: 60:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2115 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 762..2115
: US-09-146-249A-60

```

```

Query Match          6.3%; Score 37.6; DB 3; Length 2115;
Best Local Similarity 63.0%; Pred No. 0.23;
Matches 58; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 212 ctctcgcgctgtgcatacccttgatacacccttglttcaagcgccagccaaagcagccg 271
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 577 CTGCCCCCGCCGCCGCAACCTTGCGCCAGCCCGACCCCAAGTGTTCAGCGCGCGCGCGCG 636

QY 272 cgcctctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 303
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 637 CCCCCCTGCGCGCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCC 668

RESULT 6
US-08-206-188B-60
; Sequence 60, Application US/08206188B
; Patent No. 6100025

```

GENERAL INFORMATION:
APPLICANT: Wigler, Michael H.
APPLICANT: Colicelli, John J.
TITLE OF INVENTION: Cloning by Complementation and Related
TITLE OF INVENTION: Processes
NUMBER OF SEQUENCES: 84
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/206,188B
FILING DATE: 01-MAR-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/511,715
FILING DATE: 20-APR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36107
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 60:
SEQUENCE CHARACTERISTICS:
LENGTH: 2115 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 762..2115
US-08-206-188B-60

Query Match 6.3%; Score 37.6; DB 3; Length 2115;
Best Local Similarity 63.0%; Pred. No. 0.23;
Matches 58; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 212 cttctgctgtgtatccctctgataccctctgttcagcgccagcaagcgcg 271
DB 577 cttctgctgtgtatccctctgataccctctgttcagcgccagcaagcgcg 636
QY 272 cgcctcctcgcgcgcgatcgctgagcg 303
DB 637 ccccccttcgccccgccccgccccgccccg 668

US-09-121-321-15
Sequence 15, Application US/09121321
Patent No. 6090783
GENERAL INFORMATION:
APPLICANT: Saiga, Akihiko
APPLICANT: Orita, Satoshi
APPLICANT: Igarashi, Hisanaga
APPLICANT: Okumura, Kouichi
APPLICANT: Sakaguchi, Gaku
TITLE OF INVENTION: DNA MOLECULE RELATING TO SUPPRESSION OF
TITLE OF INVENTION: GENE EXPRESSION AND NOVEL PROTEIN
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: FISH & NEAVE

STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10020
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/121,321
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/933,803
FILING DATE: 19-SEP-1997
ATTORNEY/AGENT INFORMATION:
NAME: Haley, James F.
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: SHGN-12CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-596-9000
TELEFAX: 212-596-9090
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 3777 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 139..2151
US-09-121-321-15

Query Match 6.2%; Score 36.8; DB 3; Length 3777;
Best Local Similarity 69.4%; Pred. No. 0.46;
Matches 50; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 231 ccctgatacccccctgttcagcgccagcaagcgcgccctcccgcgcgcg 290
DB 1355 cccccgacacgccccgccccgccccgccccgccccgccccgccccg 1414
QY 291 atcgagtgagcc 302
DB 1415 AGCCCCAGAGCC 1426

US-08-933-803A-15
Sequence 15, Application US/08933803A
Patent No. 6218522
GENERAL INFORMATION:
APPLICANT: Saiga, Akihiko
APPLICANT: Orita, Satoshi
APPLICANT: Igarashi, Hisanaga
APPLICANT: Okumura, Kouichi
APPLICANT: Sakaguchi, Gaku
TITLE OF INVENTION: DNA MOLECULE RELATING TO SUPPRESSION OF
TITLE OF INVENTION: GENE EXPRESSION AND NOVEL PROTEIN
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: FISH & NEAVE
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10020
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible


```
RESULT 11
US-07-885-970A-3
; Sequence 3, Application US/07885970A
; Patent No. 5495070
; GENERAL INFORMATION:
; APPLICANT: John, Maliyakal E.
; TITLE OF INVENTION: GENETICALLY ENGINEERING COTTON
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nicholas J. Seay, Quarles & Brady
; STREET: P.O. Box 2113, First Wisconsin Plaza
; CITY: Madison
; STATE: Wisconsin
; COUNTRY: USA
; ZIP: 53701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/885,970A
; FILING DATE: 19920518
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/617,239
; FILING DATE: 21-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/253,243
; FILING DATE: 04-OCT-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27,386
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 283-2478
; TELEFAX: (608) 251-5139
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 913 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Gossypium hirsutum
; STRAIN: Coker 312
; DEVELOPMENTAL STAGE: 15 day old fiber cells
; TISSUE TYPE: fiber cells
; IMMEDIATE SOURCE:
; LIBRARY: CREF15A1
; CLONE: H6
; US-07-885-970A-3

Query Match          6.2%; Score 36.6; DB 1; Length 913;
Best Local Similarity 58.9%; Pred. No. 0.33;
Matches 63; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 198 tcataacatagactctgcgcgtcgtcatccctgcataccccctggttcaagcgc 257
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 320 TCACCCAGCTTCTCCCTCCCTGCACTCCACACCACTGCTTCTCCCTGCAACTCC 379

QY 258 agccaagcagccgcgcctcctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 304
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 380 ACCTCCAGCTTCTCCACCTCTGCACTCTCCACCTCCACCAAGCTTCTCCAC 426

RESULT 12
US-08-298-687A-3
; Sequence 3, Application US/08298687A
```

```
; Patent No. 5521078
; GENERAL INFORMATION:
; APPLICANT: John, Maliyakal E.
; TITLE OF INVENTION: GENETICALLY ENGINEERING COTTON
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nicholas J. Seay, Quarles & Brady
; STREET: P.O. Box 2113, First Wisconsin Plaza
; CITY: Madison
; STATE: Wisconsin
; COUNTRY: USA
; ZIP: 53701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/298,687A
; FILING DATE:
; CLASSIFICATION: 800
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/617,239
; FILING DATE: 21-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/253,243
; FILING DATE: 04-OCT-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27,386
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 283-2478
; TELEFAX: (608) 251-5139
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 913 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Gossypium hirsutum
; STRAIN: Coker 312
; DEVELOPMENTAL STAGE: 15 day old fiber cells
; TISSUE TYPE: fiber cells
; IMMEDIATE SOURCE:
; LIBRARY: CREF15A1
; CLONE: H6
; US-08-298-687A-3

Query Match          6.2%; Score 36.6; DB 1; Length 913;
Best Local Similarity 58.9%; Pred. No. 0.33;
Matches 63; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 198 tcataacatagactctgcgcgtcgtcatccctgcataccccctggttcaagcgc 257
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 320 TCACCCAGCTTCTCCCTCCCTGCACTCCACACCACTGCTTCTCCCTGCAACTCC 379

QY 258 agccaagcagccgcgcctcctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 304
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 380 ACCTCCAGCTTCTCCACCTCTGCACTCTCCACCTCCACCAAGCTTCTCCAC 426

RESULT 13
US-08-530-797-2
; Sequence 2, Application US/08530797
; Patent No. 5597718
; GENERAL INFORMATION:
; APPLICANT: John, Maliyakal E.
```



```

Query Match      6.2%: Score 36.6; DB 1; Length 913;
Best Local Similarity 58.9%: Pred. No. 0.33;
Matches 63; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 198 tcataccaatcgcgcattctgcgcgtcgtcgtcatccctcgtcatcaaccctggttcagcgc 257
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 320 TCCACCCAGTTCTCTCTCTCTCTCTCGCAACTTCACACACTCTGTTCTCTCTCTCGCAACTTC 379

QY 258 agccaagcagcgcgcctcctctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 304
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 380 ACCTCCACCTTCTCCACCTCTCTCGCACACTCTCCACACACTCTCTCCAC 426

RESULT 14
US-08-298-829-3
; Sequence 3, Application US/08298829
; Patent No. 5620882
; GENERAL INFORMATION:
; APPLICANT: John, Malyakal E.
; TITLE OF INVENTION: GENETICALLY ENGINEERING COTTON
; TITLE OF INVENTION: PLANTS FOR ALTERED FIBER
; NUMBER OF SEQUENCES: 33

```

```

Query Match          6.2%: Score 36.6; DB 1; Length 913;
Best Local Similarity 58.9%: Pred. No. 0.33;
Matches 63; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

OY 198  tcatccaatacgaattctgcgcgcgtgcgaatccctgataacccctgtttcagcgc 257
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 320  ttcaccacagttctcctcctcctcctgcaacttcaccacacttcttctcctccgcaacttc 379

OY 258  agccaaggcagccgcgcctcctcctgcgcgcgcgcgcgcgcgcgcgcgcgcgc 304
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 380  accctcagcttctccaccctcctgcacacacacacacacacacacacacacacac 426

RESULT 15
US-08-787-335-2
; Sequence 2, Application US/08787335
; Patent No. 5981834
; GENERAL INFORMATION:
; APPLICANT: John, Malyakal E.
; APPLICANT: Umbeck, Paul F.
; APPLICANT: Brill, Winston J.
; TITLE OF INVENTION: GENETICALLY ENGINEERED COTTON PLANTS

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2001, 17:52:08 : Search time 6407.28 Seconds
(without alignments)
1369.914 Million cell updates/sec

Title: US-09-138-735-1_COPY_1232_1825
Perfect score: 594
Sequence: 1 caggtacagcgtacgctt.....aaacagctctcgtgagcata 594

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 14155048 seqs, 7388405095 residues

Total number of hits satisfying chosen parameters: 28310096

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Pending_Patents_NA_Main:*

1: /cgnl_7/ptodata/1/pna/PCYUS.COMB.seq:*
2: /cgnl_7/ptodata/1/pna/US086.COMB.seq:*
3: /cgnl_7/ptodata/1/pna/US07.COMB.seq:*
4: /cgnl_7/ptodata/1/pna/US080.COMB.seq:*
5: /cgnl_7/ptodata/1/pna/US081.COMB.seq:*
6: /cgnl_7/ptodata/1/pna/US082.COMB.seq:*
7: /cgnl_7/ptodata/1/pna/US083.COMB.seq:*
8: /cgnl_7/ptodata/1/pna/US084.COMB.seq:*
9: /cgnl_7/ptodata/1/pna/US085.COMB.seq:*
10: /cgnl_7/ptodata/1/pna/US086.COMB.seq:*
11: /cgnl_7/ptodata/1/pna/US087.COMB.seq:*
12: /cgnl_7/ptodata/1/pna/US088.COMB.seq:*
13: /cgnl_7/ptodata/1/pna/US089.COMB.seq:*
14: /cgnl_7/ptodata/1/pna/US090.COMB.seq:*
15: /cgnl_7/ptodata/1/pna/US091.COMB.seq:*
16: /cgnl_7/ptodata/1/pna/US092.COMB.seq:*
17: /cgnl_7/ptodata/1/pna/US093.COMB.seq:*
18: /cgnl_7/ptodata/1/pna/US094.COMB.seq:*
19: /cgnl_7/ptodata/1/pna/US095.COMB.seq:*
20: /cgnl_7/ptodata/1/pna/US095.COMB.seq:*
21: /cgnl_7/ptodata/1/pna/US095.COMB.seq:*
22: /cgnl_7/ptodata/1/pna/US095.COMB.seq:*
23: /cgnl_7/ptodata/1/pna/US096.COMB.seq:*
24: /cgnl_7/ptodata/1/pna/US096.COMB.seq:*
25: /cgnl_7/ptodata/1/pna/US096.COMB.seq:*
26: /cgnl_7/ptodata/1/pna/US096.COMB.seq:*
27: /cgnl_7/ptodata/1/pna/US096.COMB.seq:*
28: /cgnl_7/ptodata/1/pna/US097A.COMB.seq:*
29: /cgnl_7/ptodata/1/pna/US097B.COMB.seq:*
30: /cgnl_7/ptodata/1/pna/US097C.COMB.seq:*
31: /cgnl_7/ptodata/1/pna/US098.COMB.seq:*
32: /cgnl_7/ptodata/1/pna/US098.COMB.seq:*
33: /cgnl_7/ptodata/1/pna/US6000.COMB.seq:*
34: /cgnl_7/ptodata/1/pna/US6002.COMB.seq:*
35: /cgnl_7/ptodata/1/pna/US6004.COMB.seq:*
36: /cgnl_7/ptodata/1/pna/US6004.COMB.seq:*
37: /cgnl_7/ptodata/1/pna/US6005.COMB.seq:*
38: /cgnl_7/ptodata/1/pna/US6006.COMB.seq:*
39: /cgnl_7/ptodata/1/pna/US6007.COMB.seq:*
40: /cgnl_7/ptodata/1/pna/US6008.COMB.seq:*
41: /cgnl_7/ptodata/1/pna/US6009.COMB.seq:*
42: /cgnl_7/ptodata/1/pna/US6010.COMB.seq:*
43: /cgnl_7/ptodata/1/pna/US6011.COMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	ID	Description
1	594	100.0	3402	US-08-988-242-1	Sequence 1, Appl1
2	594	100.0	3402	US-09-138-736-1	Sequence 1, Appl1
3	46	7.7	606	US-09-699-999-5885	Sequence 5885, Ap
4	45.8	7.7	470	US-09-699-999-4102	Sequence 4102, Ap
5	44.6	7.5	599	US-09-396-887-4576	Sequence 4576, Ap
6	42.6	7.2	6257	US-09-620-392-44819	Sequence 44819, A
7	42.6	7.2	29301	US-09-702-134-5721	Sequence 5721, Ap
8	42.4	7.1	600	US-09-620-392-45296	Sequence 45296, A
9	42.2	7.1	491	US-09-197-872-57273	Sequence 57273, A
10	42.2	7.1	8953	US-09-620-392-68153	Sequence 68153, A
11	42.2	7.1	10344	US-09-620-392-65666	Sequence 65666, A
12	42.2	7.1	10348	US-09-702-134-8344	Sequence 8344, Ap
13	42	7.1	7227	US-09-168-139-2273	Sequence 2273, Ap
14	42	7.1	26276	US-09-514-000-217	Sequence 217, App
15	42	7.1	397803	US-09-739-449-937	Sequence 937, App
16	41.4	7.0	602	US-09-620-392-19446	Sequence 19436, A
17	41.2	6.9	12871	US-09-620-392-3460	Sequence 3460, Ap
18	41.2	6.9	33579	US-09-702-134-4946	Sequence 4946, Ap
19	40.8	6.9	415	US-09-207-458-150303	Sequence 150303,
20	40.8	6.9	547	US-09-207-458-149725	Sequence 149725,
21	40.8	6.9	573	US-09-553-094-5427	Sequence 5427, Ap
22	40.8	6.9	635	US-09-209-830-60545	Sequence 60545, A
23	40.8	6.9	637	US-09-207-458-151348	Sequence 151348,
24	40.8	6.9	668	US-09-207-458-151301	Sequence 151301,
25	40.6	6.8	550	US-09-197-872-44432	Sequence 44492, A
26	40.4	6.8	470	US-09-654-617-111007	Sequence 111007,
27	40.4	6.8	470	US-09-684-016-111007	Sequence 111007,
28	40.4	6.8	477	US-09-135-954-1	Sequence 1, Appl1
29	40.4	6.8	569	US-09-211-750-30151	Sequence 30151, A
30	40.4	6.8	5537	PCT-US01-04098A-2481	Sequence 2481, Ap
31	40.4	6.8	5537	US-09-496-914A-6707	Sequence 6707, Ap
32	40.4	6.8	5537	US-09-560-875A-6707	Sequence 6707, Ap
33	40.4	6.8	6248	PCT-US01-04098A-513	Sequence 513, App
34	40.2	6.8	811	US-09-654-617-192521	Sequence 392521,
35	40.2	6.8	811	US-09-684-016-192521	Sequence 392521,
36	40.2	6.8	811	US-09-684-016-192521	Sequence 392521,
37	40	6.7	6867	US-09-620-392-66151	Sequence 66151, A
38	40	6.7	7443	US-09-620-392-55872	Sequence 55872, A
39	40	6.7	8299	US-09-620-392-54330	Sequence 34330, A
40	40	6.7	97314	US-09-620-392-40125	Sequence 40125, A
41	39.8	6.7	431	US-09-702-134-10211	Sequence 10211, A
42	39.8	6.7	431	US-09-394-745-23012	Sequence 23012, A
43	39.8	6.7	431	US-09-565-306-47192	Sequence 47192, A
44	39.8	6.7	1071	US-09-654-617-260110	Sequence 260110,
45	39.8	6.7	1071	US-09-684-016-260110	Sequence 260110,
45	39.8	6.7	16482	US-09-620-392-3109	Sequence 3109, Ap

ALIGNMENTS

RESULT 1
US-08-988-242-1
Sequence 1, Application US/08988242
GENERAL INFORMATION:
APPLICANT: PARAMHOS-BACCALA, GLAUCIA
APPLICANT: LESENECHAL, MYLENE
APPLICANT: JOLIVET, MICHEL
APPLICANT: MANDRAND, BERNARD
TITLE OF INVENTION: NEW TRYPAÑOSOMA CRUZI ANTIGEN, GENE
TITLE OF INVENTION: ENCODING THEREFOR, AND METHODS OF DETECTING AND TREATING
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIVE & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: Alexandria
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/988,242
FILING DATE: 10-DEC-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: MPB 36400A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3402 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-988-242-1

Query Match 100.0%; Score 594; DB 13; Length 3402;
Best Local Similarity 100.0%; Pred. No. 1.7e-126;
Matches 594; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggtacagcgttaacggtcttctgctcaatcgtacacgcagcaggtacgtctgctgct 60
DB 1232 CAGGTACAGCGTAAACGCTTTTGGCTTCAATCGTACAGCCACGCTACGCTGCTGCT 1231

QY 61 gacatgcatcgtatgcatgcatcttcacatctccggtccctcccgcaaggaagaacagcag 120
DB 1292 GACATGCGATTCGATGAGCATCTTCATCTCCGGTCTCTCCCGACGGAAGAACACAG 1351

QY 121 ccagggccaaacatcgtgtagtgagcgagcgaggaacgggggtgtgtgtctctcgggcact 180
DB 1352 CCAAGGCCAAACATCGTAGTGCGACGCGCAACCGGGGTGTGTCTCTCGGGCAGCT 1411

QY 181 gacgcgagagtagcagctacacaaatgacactctgcgctgctgacatccctcgacatca 240
DB 1412 GACGCGGAGTAGCATCTACCAATACGACTTTCTCCGCTGCTGATCCCTGCATCA 1471

QY 241 ccccgctgttcacgcagcagcagcagcgcgcctctctgcgcgagcgatcgctgag 300
DB 1472 CCCCGCTTTCAGCGCCAGCGCAGCGCGCGCTCTCGCGCGGCGCATCGGCTGAG 1531

QY 301 ccgcagctggggagcagaatcattgtctaactagtgaatcagctgggatttaatgtacc 360
DB 1532 CCGCAGCTGGGGAGCAAGATCATGTCTAATCTAGTGAATCAGCTGGGATTAATGTACC 1591

QY 361 caaaggagcgtctcaacacttgagcgcgcgcagcagcagaggtcttaagcggtgacgtcc 420
DB 1592 CAAAGGAGCGTCTCAACACTTGGAGCCGCCACGACGAGGTCTACGGCGGTACGCTCC 1651

QY 421 acgaactaccgcccgccagcagaacagtcacatagcgagacaatgycgacactgtgacgct 480
DB 1652 ACGACTACGCCCGCCAGCGCAAGTCAATACGGGCAATGGCGCAGCTGTAGCGGCT 1711

QY 481 ggatgtgtgcagctaatatagtgtgccaagcgcgctctgctcccaagccggcgagaa 540
DB 1712 GGATGTGTGCAGCTAATATAGTGTGCGCAGCGGCTCTGCTCCACAGCGCGCGCAAA 1771

QY 541 ccaacagagaagaagaagcgtcccgagcagctggaacagcagcctcggtgagata 594
DB 1772 CCAACAGGAGAGAAAGGCTCCGCGCATGTGAAACGAGCTCCGTGGCGATA 1825

RESULT 2
US-09-138-736-1
Sequence 1, Application US/09138736
GENERAL INFORMATION:
APPLICANT: PARAMHOS-BACCALA, GLAUCIA
APPLICANT: LESENECHAL, MYLENE
APPLICANT: JOLIVET, MICHEL
TITLE OF INVENTION: NEW TRYPAÑOSOMA CRUZI ANTIGEN, AND GENE
TITLE OF INVENTION: ENCODING THE LATTER, THEIR APPLICATION TO THE DETECTION OF
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Oliff & Berridge
STREET: 700 South Washington Street, Suite 300
CITY: Alexandria
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22314
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/138,736
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/480,917
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: MPB 36400
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3402 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-138-736-1

Query Match 100.0%; Score 594; DB 15; Length 3402;
Best Local Similarity 100.0%; Pred. No. 1.7e-126;
Matches 594; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggtacagcgttaacggtcttctgctcaatcgtacacgcagcaggtacgtctgctgct 60

[illegible]

```

RESULT      8
US-60-209-830-45296
: Sequence 45296, Application US/60209830
: GENERAL INFORMATION:
:   APPLICANT:      Andersen, Scott E.
:   APPLICANT:      Castiglioni, Paolo
:   APPLICANT:      Hardeman, Kristine J.
:   APPLICANT:      Kerk, Nancy M.
:   APPLICANT:      La Rosa, Thomas J.
:   TITLE OF INVENTION:  NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
:   TITLE OF INVENTION:  PLANTS
:   FILE REFERENCE:      38-21(51934)A
:   CURRENT APPLICATION NUMBER: US/60/209,830
:   CURRENT FILING DATE:   2000-06-06
:   NUMBER OF SEQ ID NOS: 62628
:   SEQ ID NO 45296
:   LENGTH: 600
:   TYPE: DNA
:   ORGANISM: Zea mays
:   OTHER INFORMATION: Clone ID: LIB3596-038-Q6-K6-G7
US-60-209-830-45296

```

	Query March	7.1%;	Score 42.4;	DB 52;	Length 600;
	Best Local Similarity	60.3%;	Pred. No. 8.4;		
	Matches 70;	Conservative	0;	Mismatches 46;	Indels 0;
	Gaps	0;			
QY	178	actgacgcggcagatagatcatcaatacgaattctgcgcgtcgatccctgca	237		
Dd	399	acaagcccacagcagtcgcgtctgcacctctctaagcgcgccaatccgcctctctcog	458		
QY	238	tcaacccttgttcacgcgcacgaaggcagcgctctctcgcgcgcgcgatc	293		
Dd	459	tcaagcgccttctcgcgcgcctccaecgggggcgcgcgcgcgcgaagtgatc	514		

```

RESULT      9
US-60-197-872-57273/C
: Sequence 57273, Application US/60197872
: GENERAL INFORMATION:
: APPLICANT: Bougri1, Oleags
: APPLICANT: Byrum, Joseph R.
: APPLICANT: De La Pena, Robert C.
: APPLICANT: La Rosa, Thomas J.
: APPLICANT: Shukla, Hridayabhiranjan
: TITLE OF INVENTION: Nucleic acid molecules and other molecules associated with
: TITLE OF INVENTION: plants
: FILE REFERENCE: 38-21(51892)A
: CURRENT APPLICATION NUMBER: US/60/197,872
: CURRENT FILING DATE: 2000-04-19
: NUMBER OF SEQ ID NOS: 76255
: SEQ ID NO 57273
:
: LENGTH: 491
: TYPE: DNA
: ORGANISM: Oryza sativa nipponbare
: OTHER INFORMATION: Clone ID: jC-osfILIB3479034a01b1
US-60-197-872-57273

```

Query Match	7.18;	Score 42.2;	DB 51;	Length 491;
-------------	-------	-------------	--------	-------------

	Best Local Similarity	53.3%	Pred. No. 9.2;	
	Matches 89;	Conservative 0;	Mismatches 78;	Indels 0; Gaps 0;
OY	142	gtggagacgagcgaacccggggtgtgtgtccttgagcactgaacgagcgagtagacagat	201	
Db	371	GGGGAGCGTCAGCGCGCGCTTGTGGTTCGAGGACTGACGAGAACGTGTTCCAGTCC	312	
OY	202	accatacgaactctgcgcgtctgcatccctccgtgataccccctglttcagcagcagcc	261	
Db	311	GCCGCTTGCGCACTTCGTAACCTGCTCAGCCCCGCCGCCGCCGCACTCTCGCGCAGCG	252	
OY	262	aagcagacgcgcgcctcctcgtccgagcgagcagctcgcgtctgaacgcgacagt	308	
Db	251	GCTGGCGCTGCGCTTCTTGCACTCTCTGCTGGCGCTGCGCACTGATGT	205	

```

RESULT 10
US-09-620-392-68153/c
: Sequence 68153, Application US/09620392
: GENERAL INFORMATION:
: APPLICANT: Boukharov, Andrey A.
: APPLICANT: Kovalic, David K.
: APPLICANT: Liu, Jingdong
: APPLICANT: McIninch, James
: TITLE OF INVENTION: Plant Genome Sequence And Uses Thereof
: FILE REFERENCE: 38-21(51237)E
: CURRENT APPLICATION NUMBER: US/09/620,392
: CURRENT FILING DATE: 2000-07-19
: NUMBER OF SEQ ID NOS: 69652
: SEQ ID NO 68153
: LENGTH: 8953
: TYPE: DNA
: ORGANISM: Oryza sativa
US-09-620-392-68153

```

Query Match	7.1%	Score 42.2	DB 24	Length 8953
Best Local Similarity	51.9%	Pred. No. 11		
Matches	95	Conservative	0	Mismatches 88
				Indels 0
				Gaps 0
QY 133	acatcgtagtctgacgacgagaaacggggtgtgtgtcctcgcgagcactgaacgagcgaagt	192		
DB 1548	ACACCCGGGCGCTGGCGCTTCACACACCGTCGCGTTCGGCGCGGCAGATCCCGGCGCGGT	1489		
QY 193	agcaatcatataccaataagactctctgcgtctgtcatccctcgatataccccctgtttaa	252		
DB 1488	ACGAATCTCGACACCCCGGGTTCGCTTCACCCCTCTCCACGACACCGGACACCCCGGCT	1429		
QY 253	ggcgcaagcaaaagcagcgacgcctcctctgcgcgagcgatcgtgacgcagcctgtggg	312		
DB 1428	CCTCATTCACAGACGCGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	1369		
QY 313	agc 315			
DB 1368	ATC 1366			

```

RESULT 11
US-09-620-392-65686
; Sequence 65686, Application US/09620392
; GENERAL INFORMATION:
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Kowalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: McIninch, James
; TITLE OF INVENTION: Plant Genome Sequence And Uses Thereof
; FILE REFERENCE: 38-21(51237)E
; CURRENT APPLICATION NUMBER: US/09/620,392
; CURRENT FILING DATE: 2000-07-19
; NUMBER OF SEQ ID NOS: 69652
; SEQ ID NO 65686
; LENGTH: 10344
; TYPE: DNA

```


THIS PAGE BLANK (USPTO)

[illegible]

```

RESULT      2
US-09-803-110-217
: Sequence 217: Application US//09803110
: GENERAL INFORMATION:
: APPLICANT: Hinkle, Gregory J.
: APPLICANT: Slater, Steven C.
: TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
: FILE REFERENCE: 38-10(115490)D
: CURRENT APPLICATION NUMBER: US//09/803,110
: CURRENT FILING DATE: 2001-03-12
: PRIOR APPLICATION NUMBER: US 09/739,449
: PRIOR FILING DATE: 2000-12-19
: PRIOR APPLICATION NUMBER: US 09/514,000
: PRIOR FILING DATE: 2000-02-23
: PRIOR APPLICATION NUMBER: US 60/168,139
: PRIOR FILING DATE: 1999-12-01
: NUMBER OF SEQ ID NOS: 13351
: SEQ ID NO 217
: LENGTH: 397803
: TYPE: DNA
: ORGANISM: Agrobacterium tumefaciens
: FEATURE:
: NAME/KEY: unsure
: LOCATION: (1)..(397803)
: OTHER INFORMATION: unsure at all n locations
US-09-803-110-217

```

```

Query Match 7.1%; Score 42; DB 6; Length 397803;
Best Local Similarity 60.5%; Pred. No. 1.8;
Matches 69; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

Oy 183 cgcgcgcagtagcagtcataccaataägcactctgcgcgtctgcatccctgcatacc 242
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 66682 cccgcgcacatctgcctccaacttacaactccgcgcgcgcgcagccctccagcacc 66741

Oy 243 ccccttccagcgcgcagcacaagcgcgcgcctctccgcgcgcgcagtcgc 296
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 66742 ggaagtctgcacgcgcgcagcgtatctctgcctctccagcgcgcgtccaccgc 66795

RESULT 3
US-09-684-1/c
: Sequence 1, Application US/09684843A
: GENERAL INFORMATION:
: APPLICANT: Ranum et al.
: TITLE OF INVENTION: SCAT GENE AND METHODS OF USE
: FILE REFERENCE: Regents of the University of Minnesota
: CURRENT APPLICATION NUMBER: US/09/684,843A
: CURRENT FILING DATE: 2000-10-06
: PRIOR APPLICATION NUMBER: 60/056,170

```

```

;
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 09/135,994
;
; PRIOR FILING DATE: 1998-08-18
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
;
; SEQ ID NO 1
;
; LENGTH: 477
;
; TYPE: DNA
;
; ORGANISM: Homo sapiens
;
US-09-684-843A-1

```

[illegible]

```

RESULT      4
US-09-922-480-3
Sequence 3, Application US/09922480
GENERAL INFORMATION:
APPLICANT: Sheppard, Paul O.
APPLICANT: Adler, David A.
TITLE OF INVENTION: SECRETED SALIVARY ZSIG63 POLYPEPTIDE
FILE REFERENCE: 97-71
CURRENT APPLICATION NUMBER: US/09/922,480
CURRENT FILING DATE: 2001-08-03
PRIOR APPLICATION NUMBER: US 60/124,820
PRIOR FILING DATE: 1999-03-17
NUMBER OF SEQ ID NOS: 9
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 3
LENGTH: 657
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Degenerate polynucleotide sequence for zsig63
NAME/KEY: misc-feature
LOCATION: (1)...(657)
OTHER INFORMATION: n = A,T,C or G
US-09-922-480-3

```

	Query Match	6.5%	Score 38.8;	DB 7;	Length 657;
	Best Local Similarity	31.4%;	Pred. No. 4.7;	Mismatches 89;	Conservative 17; Indels 0; Gaps 0;
Oy	157	cggaggtgtgtgctcctcgacatgaaggcgagtagaatcaataacgatctct	216	:	:
Db	340	ccnwmnmnttlytlrwsngcngcngcnncngcngcncncnatlbgcngcngarcn	399	:	:
Oy	217	ggcgtgtcatgcccccgcgatcacccccgttttcagcgcgcagccaagcgacgcct	276	:	:
Db	400	ggcngcngcngcncnyhtacngcnaecncngtngcngcngarcngcngcngcngcncn	459	:	:
Oy	277	ccctgcggcgcgcatcgtgccttgagccgcacgttggggagaagaattatgtcctaattagt	336	:	:
Db	460	gtngcngcngarcngcngcngcngcngcncmtnlgcngcngarcngcngcngarcncn	519	:	:
Oy	337	aatcagctcgtgatatagtcacccaagagcgctgcacgactgtgagccgcgcgcacy	396	:	:

APPLICANT: Glenn, Matthew
 APPLICANT: Norris, Michael G.
 TITLE OF INVENTION: Compositions isolated from forage
 TITLE OF INVENTION: grasses and methods for their use
 FILE REFERENCE: 1056P
 CURRENT APPLICATION NUMBER: US/60/253,654
 CURRENT FILING DATE: 2000-11-28

```

: NUMBER OF SEQ ID NOS: 32581
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO: 1232
: LENGTH: 405
: TYPE: DNA
: ORGANISM: Lolium perenne
: US-60-253-654-1232

```

Query Match:	6.48:	Score 37.8:	DB 8:	Length 405:
Best Local Similarity:	45.28:	Pred. No. 7.6:		
Matches 126:	Conservative	0:	Mismatches 147:	Indels 0:
				Gaps 0:

[illegible]

RESULT 12
US-60-255-592-1232

```

; Sequence 1232, Application US/60255592
; GENERAL INFORMATION:

```

```

; APPLICANT: Glenn, Matthew
; APPLICANT: Norriss, Michael G.

```

```

; TITLE OF INVENTION: Compositions isolated from forage
; TITLE OF INVENTION: grasses and methods for their use
; TITLE OF INVENTION:

```

```

; FILE REFERENCE: 1056P2
; CURRENT APPLICATION NUMBER: US/60/255,592

```

```

; CURRENT FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 32581

```

```
; SOFTWARE: FastSeq for Windows Version 4.0.0
; SEQ ID NO 1232
```

```

; LENGTH: 405
; TYPE: DNA
;

```

US-60-255-592-1232

Query Match	6.4%	Score 37.8	DB 8	Length 405
Best Local Similarity	46.2%	Pred. No. 7.6		
Matches 126; Conservative	0	Mismatches 147	Indels 0	Gaps 0

[illegible]

QY	393	cacgacgaggtctacggcggtgacgtccacgac	425
Db	323	cgcgacgacgctacgcgccacgaccccgcc	355

RESULT 13
US-60-253-654-18403
; Sequence 18403, Application US/60253654

; GENERAL INFORMATION:
; APPLICANT: Glenn, Matthew

APPLICANT: NORRIS, Michael G.
TITLE OF INVENTION: Composition

FILE REFERENCE: 1056P
TITLE OF INVENTION: grasses and methods for their use

```

; CURRENT APPLICATION NUMBER: US/60/253,654
; CURRENT FILING DATE: 2000-11-28

```

```

; NUMBER OF SEQ ID NOS: 32581
;
; SOFTWARE: FastSeq for Windows Version 4.0
;

```

```

; SEQ ID NO 18403
; LENGTH: 578

```

```

; TYPE: DNA
; ORGANISM: Festuca arundinaceae
;

```

05-60-253-654-18403

Query Match	6.4%	Score 37.8;	DB 8;	Length 578;
Best Local Similarity	64.0%	Pred No. 7.9;		
Matches 57; Conservative	0;	Mismatches 32;	Indels 0;	Gaps 0;

QY 209 cgactctgcgcgtgtcgtcatccctgtcatcaccctgtttcagcgccagccaaagcgag 268
 |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 330 cgacgtgtccgcgcgtctgtcgcaacacgcgcacgcgtctcgagcgagcagcccccgcag 389

QY 269 ccgcgctcctctgcgcgcgcgatcgct 297

Db 390 ccacgacgcgacggygcagggcggt 418

RESULT 14
US-60-255-592-18403

; sequence 18403, Application US/6025559,
; GENERAL INFORMATION:
; GENERAL INFORMATION:

```

; APPLICANT: Glenn, Matthew
; APPLICANT: Norris, Michael G.

```

```

; TITLE OF INVENTION: Compositions isolated from forage
; TITLE OF INVENTION: grasses and methods for their use
;

```

```

; FILE REFERENCE: 1056P2
; CURRENT APPLICATION NUMBER: US/60/255,593

```

```

; CURRENT FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 32581

```

```
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 18403
```

```

; LENGTH: 5/8
; TYPE: DNA
;

```

US-60-255-592-18403

Query Match	6.4%	Score 37.8	DB 8	Length 578
Best Local Similarity	64.0%	Pred No. 7.9		
Matches 57	Conservative	0	Mismatches 32	Indels 0
			Gaps	0

Qy 209 cgactctgcgcgtctgcataccctgcataccctgtttcagcgcagccaagcag 265
 |||| | |||| | | ||| | | ||| | ||| |||| ||||
 Db 330 cgacgtgctcgcgcgtctgcataccacgcgcacgcgtctcagcgcgcagccccgcag 385

DY 269 ccgcgcctcctgccgcgccgatcggt 297

Db 390 ccaacgagcgagcgggcgagggcggt 418

RESULT 15

